OM protein - protein search, using sw model

Run on: January 30, 2004, 00:13:12 ; Search time 8.97183 Seconds
(without alignments)
225.098 Million cell updates/sec

Title: US-09-461-684C-1
Perfect score: 109
Sequence: 1 CKKKKKKKKKKKKKKKKKKKK 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	100	91.7	215	2	I52523	nucleoporin p62 ho
2	100	91.7	380	2	T46395	hypothetical prote
3	100	91.7	517	2	T49173	hypothetical prote
4	93	85.3	166	2	T18513	hypothetical prote
5	90	82.6	483	2	F71619	hypothetical prote
6	87	79.8	107	2	C86477	protein F15O4.29 [
7	84	77.1	441	2	A48455	acidic phosphoprot
8	81	74.3	4550	2	T18440	hypothetical prote
9	80	73.4	784	2	T18452	hypothetical prote
10	78	71.6	529	2	T50609	hypothetical prote
11	78	71.6	560	2	T06377	SAR DNA-binding pr
12	78	71.6	1560	2	T42727	proliferation pote
13	77	70.6	229	2	JC7219	nuclear protein SR
14	77	70.6	266	2	A86288	protein F9L1.30 [i
15	77	70.6	3724	2	T18427	hypothetical prote
16	76	69.7	511	2	S58322	nucleolar protein
17	76	69.7	812	2	S43604	R07E5.1 protein (c
18	75	68.8	474	2	T38485	centromere/microtu
19	75	68.8	504	2	S48550	hypothetical prote
20	74	67.9	191	2	T23594	hypothetical prote
21	74	67.9	198	2	D97301	probable membrane
22	74	67.9	414	2	F64386	hypothetical prote
23	74	67.9	483	2	S41853	centromere/microtu
24	73	67.0	212	2	T49559	related to proline
25	72.5	66.5	683	2	T00872	probable protein k
26	72	66.1	404	2	A75192	hypothetical prote
27	72	66.1	560	2	T29586	hypothetical prote
28	72	66.1	678	2	A54514	glutamic acid-rich
29	72	66.1	1701	2	T09127	probable erythrocy
30	72	66.1	4981	2	T18489	hypothetical prote
31	71	65.1	683	2	T34103	hypothetical prote
32	71	65.1	686	1	A44842	cGMP-gated ion cha
33	71	65.1	690	2	A42161	cGMP-gated cation
34	70.5	64.7	163	2	T42696	hypothetical prote
35	70	64.2	222	2	T28919	hypothetical prote
36	70	64.2	233	2	S55165	hypothetical prote
37	70	64.2	291	2	T48617	hypothetical prote
38	70	64.2	451	2	T16418	hypothetical prote
39	70	64.2	497	2	T29814	hypothetical prote
40	70	64.2	508	2	E71620	hypothetical prote
41	70	64.2	688	2	B42161	cGMP-gated cation
42	70	64.2	2013	2	C71610	probable membrane
43	69.5	63.8	2231	2	S53416	SEN1 protein - yea
44	69	63.3	142	2	S54481	hypothetical prote
45	69	63.3	167	2	S38112	hypothetical prote

ALIGNMENTS

RESULT 1
I52523
nucleoporin p62 homolog - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
C;Accession: I52523
R;Wang, Z.Q.; Akmal, K.M.; Kim, K.H.
Biol. Reprod. 51, 1022-1030, 1994
A;Title: An unusual nucleoporin-related messenger ribonucleic acid is present in the gern
A;Reference number: I52523; MUID:95151924; PMID:7849178
A;Accession: I52523
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-215 <RES>
A;Cross-references: GB:S75997; NID:g913245; PIDN:AAB33384.1; PID:g913246
A;Experimental source: testis

```
  Query Match             91.7%;  Score 100;  DB 2;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 0.0017;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db         35 KKKKKKKKKKKKKKKKKKKK 54
```

RESULT 2
T46395
hypothetical protein DKFZp434I1120.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46395
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46395
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-380 <AAA>
A;Cross-references: EMBL:AL137556
A;Experimental source: adult testis; clone DKFZp434I1120
C;Genetics:
A;Note: DKFZp434I1120.1

```
  Query Match             91.7%;  Score 100;  DB 2;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 0.0025;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        355 KKKKKKKKKKKKKKKKKKKK 374
```

```
RESULT 3
T49173
hypothetical protein T20N10.250 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
C;Accession: T49173
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25017
A;Accession: T49173
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-517 <DAN>
A;Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
A;Experimental source: cultivar Columbia; BAC clone T20N10
C;Genetics:
A;Gene: ATSP:T20N10.250
A;Map position: 3
A;Introns: 312/3; 359/3; 444/3
C;Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

  Query Match             91.7%;  Score 100;  DB 2;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 0.003;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        444 KKKKKKKKKKKKKKKKKKKK 463

RESULT 4
T18513
hypothetical protein C0845c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T18513
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18513
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-166 <LAW>
A;Cross-references: EMBL:Z98551; PIDN:CAB11123.2
C;Genetics:
A;Map position: 3
A;Introns: 19/1
A;Note: C0845c

  Query Match             85.3%;  Score 93;  DB 2;  Length 166;
  Best Local Similarity   90.0%;  Pred. No. 0.0067;
  Matches   18;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||:||||:||||||||
Db         37 KKKKKKEKKKKRKKKKKKKK 56

RESULT 5
F71619
hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: F71619
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: F71619
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-483 <GAR>
A;Cross-references: GB:AE001382; GB:AE001362; NID:g3845130; PIDN:AAC71836.1; PID:g384513
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0235w

  Query Match             82.6%;  Score 90;  DB 2;  Length 483;
  Best Local Similarity   85.0%;  Pred. No. 0.026;
  Matches   17;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||:|||::||||||
Db        449 KKKKKKKKRKKKERKKKKKK 468

RESULT 6
C86477
protein F15O4.29 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C86477
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86477
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-107 <STO>
A;Cross-references: GB:AE005172; NID:g8778346; PIDN:AAF79354.1; GSPDB:GN00141
C;Genetics:
A;Gene: F15O4.29
A;Map position: 1

  Query Match             79.8%;  Score 87;  DB 2;  Length 107;
  Best Local Similarity   94.4%;  Pred. No. 0.019;
  Matches   17;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKK 19
              |||||||||||||||||:
Db         29 KKKKKKKKKKKKKKKKKR 46

RESULT 7
A48455
acidic phosphoprotein PcEMA1q - Plasmodium chabaudi
C;Species: Plasmodium chabaudi
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C;Accession: A48455
R;Deleersnijder, W.; Prasomsitti, P.; Tungpradubkul, S.; Hendrix, D.; Hamers-Casterman, C
Mol. Biochem. Parasitol. 56, 59-68, 1992
A;Title: Structure of a Plasmodium chabaudi acidic phosphoprotein that is associated wit
A;Reference number: A48455; MUID:93116806; PMID:1475002
A;Accession: A48455
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-441 <DEL>
A;Cross-references: GB:M95789; NID:g160602; PID:g160603
A;Experimental source: IP-PC1/C
A;Note: sequence extracted from NCBI backbone (NCBIN:121415, NCBIP:121416)
C;Keywords: phosphoprotein

  Query Match             77.1%;  Score 84;  DB 2;  Length 441;
  Best Local Similarity   80.0%;  Pred. No. 0.089;
  Matches   16;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||:||||||:|:|||:
Db        394 KKKKKKEKKKKKKEKEKKKE 413
```

RESULT 8
T18440
hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18440
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18440
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4550 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB11121.1
C;Genetics:
A;Map position: 3
A;Note: C0425w

```
  Query Match             74.3%;  Score 81;  DB 2;  Length 4550;
  Best Local Similarity   44.2%;  Pred. No. 0.77;
  Matches   19;  Conservative    2;  Mismatches    0;  Indels   22;  Gaps    1;

Qy          1 CKKK----------------------KKKKKKKKKKKKKKKK 21
              ||||                      |||:||||||||||:||
Db        132 CKKKNYFYNIKNKYFNEKFYQTNNNIKKKRKKKKKKKKKKEKK 174
```

RESULT 9
T18452
hypothetical protein C0560c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18452
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z18937
A;Accession: T18452
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-784 <LAW>
A;Cross-references: EMBL:AL008970; NID:e1407852; PID:e1332545; PIDN:CAA15594.1
C;Genetics:
A;Map position: 3
A;Note: C0560c

```
  Query Match             73.4%;  Score 80;  DB 2;  Length 784;
  Best Local Similarity   75.0%;  Pred. No. 0.31;
  Matches   15;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              :||||:||||:||||:|||:
Db        360 EKKKKQKKKKQKKKKQKKKQ 379
```

RESULT 10
T50609
hypothetical protein DKFZp761B2423.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T50609
R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, June 2000
A;Reference number: Z25143
A;Accession: T50609
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-529 <AAA>
A;Cross-references: EMBL:AL359564
A;Experimental source: adult amygdala; clone DKFZp761B2423
C;Genetics:
A;Note: DKFZp761B2423.1

```
  Query Match             71.6%;  Score 78;  DB 2;  Length 529;
  Best Local Similarity   80.0%;  Pred. No. 0.37;
  Matches   16;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              |||||||:|||  ||  ||  ||
Db        464 KKKKKKKEKKKHKKHKKHKK 483
```

RESULT 11
T06377
SAR DNA-binding protein-1 - garden pea
C;Species: Pisum sativum (garden pea)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 02-Jun-2000
C;Accession: T06377
R;Hatton, D.; Gray, J.C.
submitted to the EMBL Data Library, April 1998
A;Description: cDNA encoding a pea SAR DNA-binding protein that shows homology to nucleo]
A;Reference number: Z15637
A;Accession: T06377
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-560 <HAT>
A;Cross-references: EMBL:AF061962; NID:g3132695; PIDN:AAC16330.1; PID:g3132696
C;Genetics:
A;Gene: SARBP-1
C;Superfamily: garden pea SAR DNA-binding protein

```
  Query Match             71.6%;  Score 78;  DB 2;  Length 560;
  Best Local Similarity   75.0%;  Pred. No. 0.38;
  Matches   15;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              |||:||:||:|| ||:||||
Db        463 KKKEKKEKKEKKDKKEKKKK 482
```

RESULT 12
T42727
proliferation potential-related protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000
C;Accession: T42727
R;Witte, M.M.; Scott, R.E.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z22246
A;Accession: T42727
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1560 <WIT>
A;Cross-references: EMBL:U83913; NID:g3858884; PID:g3858885; PIDN:AAC72432.1
A;Experimental source: strain Balb/C
C;Genetics:
A;Gene: P2P-R
C;Function:
A;Description: involved in hnRNP association and Rb1 binding
C;Superfamily: RING finger homology
F;57-107/Domain: RING finger homology <RRN>

```
  Query Match             71.6%;  Score 78;  DB 2;  Length 1560;
  Best Local Similarity   80.0%;  Pred. No. 0.74;
  Matches   16;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              |||||||:|||  ||  ||  ||
Db       1497 KKKKKKKEKKKHKKHKKHKK 1516
```

RESULT 13

```
JC7219
nuclear protein SR-25 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C;Accession: JC7219
R;Sasahara, K.; Yamaoka, T.; Moritani, M.; Tanaka, M.; Iwahana, H.; Yoshimoto, K.; Miyag
Biochem. Biophys. Res. Commun. 269, 444-450, 2000
A;Title: Molecular cloning and expression analysis of a putative nuclear protein, SR-25.
A;Reference number: JC7219; MUID:20175222; PMID:10708573
A;Accession: JC7219
A;Molecule type: mRNA
A;Residues: 1-229 <SAS>
A;Cross-references: DDBJ:AB035383; NID:g7619895; PIDN:BAA94743.1; PID:g7619896
A;Experimental source: MIN6 cell line
C;Comment: This protein is a highly hydrophilic nuclear protein with a serine-arginine r
A splicing factors.
C;Keywords: nucleus; RNA processing

  Query Match             70.6%;  Score 77;  DB 2;  Length 229;
  Best Local Similarity   75.0%;  Pred. No. 0.27;
  Matches   15;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
                | |:||:||||||:||| ||
Db         95 KHKEKKRKKKKKKRKKKLKK 114


RESULT 14
A86288
protein F9L1.30 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: A86288
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86288
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-266 <STO>
A;Cross-references: GB:AE005172; NID:g5103832; PIDN:AAD39662.1; GSPDB:GN00141
C;Genetics:
A;Gene: F9L1.30
A;Map position: 1

  Query Match             70.6%;  Score 77;  DB 2;  Length 266;
  Best Local Similarity   70.0%;  Pred. No. 0.3;
  Matches   14;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
                |:||::| |||| ||:||||
Db        231 KRKKRRKNKKKKNKKEKKKK 250


RESULT 15
T18427
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18427
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18427
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3724 <LAW>
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1
C;Genetics:
A;Introns: 307/1; 1545/2
A;Note: C0335c

  Query Match             70.6%;  Score 77;  DB 2;  Length 3724;
  Best Local Similarity   75.0%;  Pred. No. 1.6;
  Matches   15;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CKKKKKKKKKKKKKKKKKKK 20
                |: ||| |:||| |||||||
Db       2201 CEIKKKMKRKKKMKKKKKKK 2220

Search completed: January 30, 2004, 00:26:20
Job time : 8.97183 secs
```

```
OM protein - protein search, using sw model

Run on:         January 29, 2004, 23:58:27 ; Search time 4.73239 Seconds
                                             (without alignments)
                                             208.681 Million cell updates/sec

Title:          US-09-461-684C-1
Perfect score:  109
Sequence:       1 CKKKKKKKKKKKKKKKKKKK 21

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	84	77.1	441	1	PHPA_PLACH	Q02752 plasmodium
2	76	69.7	511	1	NOP5_YEAST	Q12499 saccharomyc
3	75	68.8	474	1	CBF5_SCHPO	O14007 schizosacch
4	75	68.8	504	1	SIK1_YEAST	Q12460 saccharomyc
5	75	68.8	2404	1	SON_MOUSE	Q9qx47 mus musculu
6	75	68.8	2426	1	SON_HUMAN	P18583 homo sapien
7	74	67.9	414	1	Y694_METJA	Q58105 methanococc
8	74	67.9	474	1	CBF5_KLULA	O13473 kluyveromyc
9	74	67.9	483	1	CBF5_YEAST	P33322 saccharomyc
10	74	67.9	726	1	BRD3_HUMAN	Q15059 homo sapien
11	72	66.1	479	1	CBF5_CANAL	O43101 candida alb
12	72	66.1	678	1	GARP_PLAFF	P13816 plasmodium
13	71	65.1	351	1	CG79_HUMAN	Q9y388 homo sapien
14	71	65.1	686	1	CNG1_HUMAN	P29973 h cgmp-gate
15	70.5	64.7	1411	1	TCOF_HUMAN	Q13428 homo sapien
16	70	64.2	233	1	YJO8_YEAST	P47006 saccharomyc
17	70	64.2	599	1	HM21_HUMAN	Q9ugu5 homo sapien
18	70	64.2	683	1	CNG1_RAT	Q62927 r cgmp-gate
19	70	64.2	684	1	CNG1_MOUSE	P29974 m cgmp-gate
20	69.5	63.8	534	1	NOP5_RAT	Q9qz86 rattus norv
21	69.5	63.8	2231	1	SEN1_YEAST	Q00416 saccharomyc
22	69	63.3	142	1	YM8H_YEAST	Q03525 saccharomyc
23	69	63.3	167	1	YK20_YEAST	P36133 saccharomyc
24	69	63.3	723	1	SSRP_DROME	Q05344 drosophila
25	69	63.3	843	1	BLVR_BOVIN	Q03368 bos taurus
26	68.5	62.8	724	1	YO61_CAEEL	P34600 caenorhabdi
27	68	62.4	523	1	DBP3_YEAST	P20447 saccharomyc
28	68	62.4	1178	1	MNN4_YEAST	P36044 saccharomyc
29	67	61.5	118	1	YQ93_CAEEL	Q10032 caenorhabdi
30	67	61.5	690	1	CNG1_BOVIN	Q00194 b cgmp-gate
31	67	61.5	691	1	CNG1_CANFA	Q28279 c cgmp-gate
32	67	61.5	1002	1	IF2P_YEAST	P39730 saccharomyc
33	67	61.5	1220	1	IF2P_HUMAN	O60841 homo sapien
34	67	61.5	1362	1	BRD4_HUMAN	O60885 homo sapien
35	67	61.5	2280	1	YCF2_OENHO	Q9mef2 oenothera h
36	66	60.6	481	1	CBF5_EMENI	O43100 emericella
37	66	60.6	487	1	CBF5_ASPFU	O43102 aspergillus
38	66	60.6	667	1	YKY1_SCHPO	O13796 schizosacch
39	66	60.6	1153	1	A3D1_HUMAN	O14617 homo sapien
40	66	60.6	1240	1	YNJ1_YEAST	P53935 saccharomyc
41	65.5	60.1	508	1	NO60_DROME	O44081 drosophila
42	65.5	60.1	514	1	DKC1_HUMAN	O60832 homo sapien
43	65	59.6	217	1	KS1_HYDAT	P38978 hydra atten
44	65	59.6	271	1	YG5W_YEAST	P53335 saccharomyc
45	65	59.6	320	1	YD33_YEAST	Q12117 saccharomyc

ALIGNMENTS

```
RESULT 1
PHPA_PLACH
ID   PHPA_PLACH     STANDARD;      PRT;   441 AA.
AC   Q02752;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Acidic phosphoprotein precursor (50 kDa antigen).
GN   PCEMA1.
OS   Plasmodium chabaudi.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5825;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IP-PC1;
RX   MEDLINE=93116806; PubMed=1475002;
RA   Deleersnijder W., Prasomsitti P., Tungpradubkul S., Hendrix D.,
RA   Hamers-Casterman C., Hamers R.;
RT   "Structure of a Plasmodium chabaudi acidic phosphoprotein that is
RT   associated with the host erythrocyte membrane.";
RL   Mol. Biochem. Parasitol. 56:59-68(1992).
CC   -!- FUNCTION: DURING INFECTION, THIS PHOSPHOPROTEIN PROBABLY MODULATES
CC       THE STRUCTURE OF THE RED CELL MEMBRANE TO THE ADVANTAGE OF THE
CC       PARASITE, ALTHOUGH ITS PRECISE FUNCTION IS NOT KNOWN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ON THE
CC       CYTOPLASMIC FACE OF THE HOST ERYTHROCYTE MEMBRANE.
CC   -!- MISCELLANEOUS: ASSOCIATED WITH THE HOST RED CELL MEMBRANE
CC       THROUGHOUT THE ENTIRE ERYTHROCYTIC CYCLE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M95789; AAA29732.1; -.
DR   PIR; A48455; A48455.
KW   Phosphorylation; Signal; Antigen; Membrane; Repeat; Erythrocyte.
FT   SIGNAL        1     15       OR 24 (POTENTIAL).
FT   CHAIN        16    441       ACIDIC PHOSPHOPROTEIN.
FT   DOMAIN      186    313       16 X 8 AA TANDEM REPEATS.
FT   REPEAT      186    193       1-1.
FT   REPEAT      194    201       1-2.
FT   REPEAT      202    209       1-3.
FT   REPEAT      210    217       1-4.
FT   REPEAT      218    225       1-5.
FT   REPEAT      226    233       1-6.
FT   REPEAT      234    241       1-7.
FT   REPEAT      242    249       1-8.
FT   REPEAT      250    257       1-9.
FT   REPEAT      258    265       1-10.
FT   REPEAT      266    273       1-11.
FT   REPEAT      274    281       1-12.
FT   REPEAT      282    289       1-13.
```

```
FT   REPEAT      290    297       1-14.
FT   REPEAT      298    305       1-15.
FT   REPEAT      306    313       1-16.
FT   DOMAIN      353    370       2 X 9 AA TANDEM REPEATS.
FT   REPEAT      353    360       2-1.
FT   REPEAT      361    368       2-2.
FT   DOMAIN      371    417       LYS-RICH (BASIC).
FT   CARBOHYD     21     21       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   441 AA;  49708 MW;  DB85E83E795EE7E5 CRC64;

  Query Match             77.1%;  Score 84;  DB 1;  Length 441;
  Best Local Similarity   80.0%;  Pred. No. 0.032;
  Matches   16;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||:||||||:|:|||:
Db        394 KKKKKKEKKKKKKEKEKKKE 413

RESULT 2
NOP5_YEAST
ID   NOP5_YEAST     STANDARD;      PRT;   511 AA.
AC   Q12499;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Nucleolar protein NOP58 (Nucleolar protein NOP5).
GN   NOP58 OR NOP5 OR YOR310C OR O6108.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RA   Pearson B.M., Hernando Y., Wolf S.S., Kalogeropoulos A., Schweizer M.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=W303;
RX   MEDLINE=98298165; PubMed=9632712;
RA   Wu P., Brockenbrough J.S., Metcalfe A.C., Chen S., Aris J.P.;
RT   "Nop5p is a small nucleolar ribonucleoprotein component required for
RT   pre-18S rRNA processing in yeast.";
RL   J. Biol. Chem. 273:16453-16463(1998).
CC   -!- FUNCTION: REQUIRED FOR PRE-18S RRNA PROCESSING. MAY BIND
CC       MICROTUBULES.
CC   -!- SUBUNIT: INTERACTS WITH NOP56 AND NOP1.
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X90565; CAA62165.1; -.
DR   EMBL; Z75217; CAA99630.1; -.
DR   EMBL; AF056070; AAC39484.1; -.
DR   PIR; S58322; S58322.
DR   SGD; S0005837; NOP58.
DR   GO; GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.
DR   GO; GO:0003754; F:chaperone activity; NAS.
DR   GO; GO:0017069; F:snRNA binding activity; IDA.
DR   GO; GO:0030490; P:processing of 20S pre-rRNA; IPI.
DR   GO; GO:0006608; P:snRNP protein-nucleus import; NAS.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
```

```
KW   Ribosome biogenesis; Nuclear protein; rRNA processing.
FT   DOMAIN       441    511       ASP/GLU/LYS-RICH.
SQ   SEQUENCE   511 AA;  56956 MW;  8A2889448B2A19E2 CRC64;

  Query Match             69.7%;  Score 76;  DB 1;  Length 511;
  Best Local Similarity   70.0%;  Pred. No. 0.21;
  Matches   14;  Conservative    6;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||:||:||:||:||:||:||
Db        480 KKEKKEKKEKKEKKEKKEKK 499

RESULT 3
CBF5_SCHPO
ID   CBF5_SCHPO     STANDARD;      PRT;   474 AA.
AC   O14007;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Centromere/microtubule binding protein cbf5 (Centromere-binding factor
DE   5) (Nucleolar protein cbf5).
GN   CBF5 OR SPAC29A4.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC       CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC       CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC   -!- SIMILARITY: Contains 1 PUA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; Z97210; CAB10131.1; -.
DR   PIR; T38485; T38485.
DR   GeneDB_SPombe; SPAC29A4.04c; -.
DR   InterPro; IPR004802; Cbf5.
DR   InterPro; IPR002478; PUA.
DR   InterPro; IPR002501; TruB_N.
DR   InterPro; IPR004521; Unchar_dom_2.
DR   Pfam; PF01472; PUA; 1.
DR   Pfam; PF01509; TruB_N; 1.
DR   SMART; SM00359; PUA; 1.
DR   TIGRFAMs; TIGR00425; CBF5; 1.
DR   TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR   PROSITE; PS50890; PUA; 1.
KW   Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT   DOMAIN      271    346       PUA.
FT   DOMAIN      434    468       7 X 3 AA APPROXIMATE TANDEM REPEATS OF
FT                                K-K-E.
FT   REPEAT      443    445       1.
FT   REPEAT      450    452       2.
FT   REPEAT      454    456       3.
FT   REPEAT      457    459       4.
FT   REPEAT      460    462       5.
FT   REPEAT      463    465       6.
FT   REPEAT      466    468       7.
SQ   SEQUENCE   474 AA;  53110 MW;  B8C9896C5FAEEB71 CRC64;

  Query Match             68.8%;  Score 75;  DB 1;  Length 474;
  Best Local Similarity   73.7%;  Pred. No. 0.24;
  Matches   14;  Conservative    5;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKK 20
              ||:||:||:||:||:||||
Db        454 KKEKKEKKEKKEKKEKKKK 472


RESULT 4
SIK1_YEAST
ID   SIK1_YEAST     STANDARD;      PRT;   504 AA.
AC   Q12460;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   SIK1 protein (Nucleolar protein NOP56).
GN   SIK1 OR NOP56 OR YLR197W OR L8167.9.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / YPH1;
RX   MEDLINE=96040178; PubMed=7547500;
RA   Morin P.J., Downs J.A., Snodgrass A.M., Gilmore T.D.;
RT   "Genetic analysis of growth inhibition by GAL4-L kappa B-alpha in
RT   Saccharomyces cerevisiae.";
RL   Cell Growth Differ. 6:789-798(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313267; PubMed=9169871;
RA   Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA   Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA   Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA   Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA   Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA   Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA   Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA   Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA   Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA   Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA   Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
```

```
RL   Nature 387:87-90(1997).
RN   [3]
RP   CHARACTERIZATION, AND MUTAGENESIS.
RX   MEDLINE=98038777; PubMed=9372940;
RA   Gautier T., Berges T., Tollervey D., Hurt E.;
RT   "Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p
RT   and are required for ribosome biogenesis.";
RL   Mol. Cell. Biol. 17:7088-7098(1997).
CC   -!- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.
CC   -!- SUBUNIT: INTERACTS WITH NOP1 AND NOP58.
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U20237; AAC49066.1; -.
DR   EMBL; U14913; AAB67431.1; -.
DR   PIR; S48550; S48550.
DR   SGD; S0004187; SIK1.
DR   GO; GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.
DR   GO; GO:0030490; P:processing of 20S pre-rRNA; IPI.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Ribosome biogenesis; Nuclear protein.
FT   DOMAIN      443    504       ASP/GLU/LYS-RICH.
FT   MUTAGEN     333    333       V->A: REDUCED GROWTH RATE AT ALL
FT                                TEMPERATURES; WHEN ASSOCIATED WITH R-385.
FT   MUTAGEN     355    355       Y->C: AT 37 DEGREES, GROWTH SLOWS AFTER 6
FT                                TO 8 HOURS AND CELL DIVISION STOPS AFTER
FT                                20 HOURS.
FT   MUTAGEN     385    385       M->R: REDUCED GROWTH RATE AT ALL
FT                                TEMPERATURES; WHEN ASSOCIATED WITH A-333.
SQ   SEQUENCE   504 AA;  56864 MW;  F8522A5870EF4842 CRC64;

  Query Match             68.8%;  Score 75;  DB 1;  Length 504;
  Best Local Similarity   70.0%;  Pred. No. 0.25;
  Matches   14;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||:||:|| ||:||:||:||
Db        465 KKEKKEKKSKKEKKEKKEKK 484


RESULT 5
SON_MOUSE
ID   SON_MOUSE      STANDARD;      PRT;  2404 AA.
AC   Q9QX47; Q9CQ12; Q9CQK6; Q9QXP5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   SON protein.
GN   SON.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   STRAIN=129/Sv;
RX   MEDLINE=20408886; PubMed=10950926;
RA   Wynn S.L., Fisher R.A., Pagel C., Price M., Liu Q.Y., Khan I.M.,
RA   Zammit P., Dadrah K., Mazrani W., Kessling A., Lee J.S., Buluwela L.;
RT   "Organization and conservation of the GART/SON/DONSON locus in mouse
RT   and human genomes.";
RL   Genomics 68:57-62(2000).
RN   [2]
```

```
RP   SEQUENCE OF 1-116 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: Transcriptional repressor. Binds to the consensus DNA
CC       sequence: 5'-GA[GT]AN[CG][AG]CC-3'. Might protect cells from
CC       apoptosis. Might be involved in pre-mRNA splicing (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9QX47-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9QX47-2; Sequence=VSP_004416, VSP_004417;
CC   -!- TISSUE SPECIFICITY: Widely expressed.
CC   -!- DOMAIN: Contains 8 types of repeats which are distributed in 3
CC       regions.
CC   -!- SIMILARITY: Contains 1 G-patch domain.
CC   -!- SIMILARITY: Contains 1 DRBM (double-stranded RNA-binding) domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF193606; AAF23120.1; -.
DR   EMBL; AF193595; AAF23120.1; JOINED.
DR   EMBL; AF193596; AAF23120.1; JOINED.
DR   EMBL; AF193597; AAF23120.1; JOINED.
DR   EMBL; AF193598; AAF23120.1; JOINED.
DR   EMBL; AF193599; AAF23120.1; JOINED.
DR   EMBL; AF193600; AAF23120.1; JOINED.
DR   EMBL; AF193601; AAF23120.1; JOINED.
DR   EMBL; AF193602; AAF23120.1; JOINED.
DR   EMBL; AF193603; AAF23120.1; JOINED.
DR   EMBL; AF193604; AAF23120.1; JOINED.
DR   EMBL; AF193605; AAF23120.1; JOINED.
DR   EMBL; AF193607; AAF23121.1; -.
DR   EMBL; AK019312; BAB31659.1; -.
DR   EMBL; AK019081; BAB31536.1; -.
DR   EMBL; AK008478; BAB25691.1; -.
DR   EMBL; AK008256; BAB25562.1; -.
DR   MGD; MGI:98353; Son.
DR   GO; GO:0005515; F:protein binding activity; IPI.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR000467; G_patch.
DR   Pfam; PF00035; dsrm; 1.
DR   Pfam; PF01585; G-patch; 1.
DR   SMART; SM00443; G_patch; 1.
DR   PROSITE; PS50137; DS_RBD; 1.
DR   PROSITE; PS50174; G_PATCH; 1.
KW   RNA-binding; DNA-binding; Nuclear protein; Repeat;
KW   Alternative splicing.
FT   DOMAIN      721    850       13 X 10 AA TANDEM REPEATS OF L-A-[ST]-
FT                                [NSG]-[TS]-MDSQM.
FT   DOMAIN      867    943       11 X 7 AA TANDEM REPEATS OF [DR]-P-Y-R-
FT                                [LI][AG][QHP].
FT   DOMAIN      961   1080       14 X 6 AA REPEATS OF [ED]-R-S-M-M-S.
FT   REPEAT      961    966       1-1.
FT   REPEAT      969    974       1-2.
FT   REPEAT      976    981       1-3.
FT   REPEAT      985    990       1-4.
FT   REPEAT      993    998       1-5.
FT   REPEAT     1001   1006       1-6.
FT   REPEAT     1010   1015       1-7.
FT   REPEAT     1018   1023       1-8.
FT   REPEAT     1026   1031       1-9.
FT   REPEAT     1035   1040       1-10.
FT   REPEAT     1044   1049       1-11.
FT   REPEAT     1055   1060       1-12.
FT   REPEAT     1066   1071       1-13.
FT   REPEAT     1075   1080       1-14.
FT   DOMAIN     1101   1133       3 X 11 AA TANDEM REPATS OF P-P-L-P-P-E-E-
FT                                P-P-[TME]-[MTG].
FT   DOMAIN     1910   1979       7 X 7 AA REPEATS OF P-S-R-R-S-R-[TS].
FT   REPEAT     1910   1916       2-1.
FT   REPEAT     1938   1944       2-2.
FT   REPEAT     1945   1951       2-3.
FT   REPEAT     1952   1958       2-4.
FT   REPEAT     1959   1965       2-5.
FT   REPEAT     1966   1972       2-6.
FT   REPEAT     1973   1979       2-7 (APPROXIMATE).
FT   DOMAIN     1919   1990       2 X 19 AA REPEATS OF P-S-R-R-R-R-S-R-S-V-
FT                                V-R-R-R-S-F-S-I-S.
FT   REPEAT     1919   1937       3-1.
FT   REPEAT     1980   1990       3-2 (APPROXIMATE).
FT   DOMAIN     1991   2017       3 X TANDEM REPEATS OF [ST]-P-[VLI]-R-
FT                                [RL]-[RK]-[RF]-S-R.
FT   DOMAIN     2283   2329       G-PATCH.
FT   DOMAIN     2349   2404       DRBM.
FT   VARSPLIC   2086   2086       K -> F (in isoform 2).
FT                                /FTId=VSP_004416.
FT   VARSPLIC   2087   2404       Missing (In isoform 2).
FT                                /FTId=VSP_004417.
SQ   SEQUENCE   2404 AA;  261428 MW;  648BF28ED3FC01D9 CRC64;

  Query Match             68.8%;  Score 75;  DB 1;  Length 2404;
  Best Local Similarity   75.0%;  Pred. No. 0.85;
  Matches   15;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              || || |  |||||||:|:||
Db        109 KKHKKHKNKKKKKKKEKEKK 128


RESULT 6
SON_HUMAN
ID   SON_HUMAN      STANDARD;      PRT;  2426 AA.
AC   P18583; O14487; O95981; Q14120; Q9H7B1; Q9P070; Q9P072; Q9UKP9;
AC   Q9UPY0;
DT   01-NOV-1990 (Rel. 16, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   SON protein (SON3) (Negative regulatory element-binding protein) (NRE-
DE   binding protein) (DBP-5) (Bax antagonist selected in saccharomyces 1)
DE   (BASS1) (Protein C21orf50).
GN   SON OR NREBP OR DBP5 OR C21ORF50 OR KIAA1019.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B; C; D; E AND F).
RX   MEDLINE=21564202; PubMed=11707072;
```

```
RA   Reymond A., Friedli M., Neergaard Henrichsen C., Chapot F.,
RA   Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,
RA   Antonarakis S.E.;
RT   "From PREDs and open reading frames to cDNA isolation: revisiting the
RT   human chromosome 21 transcription map.";
RL   Genomics 78:46-54(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM G).
RC   TISSUE=Liver;
RX   MEDLINE=21316479; PubMed=11306577;
RA   Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shiou S.-R., Lai C.-K.,
RA   Ting L.-P.;
RT   "Transcription repression of human hepatitis B virus genes by negative
RT   regulatory element-binding protein/SON.";
RL   J. Biol. Chem. 276:24059-24067(2001).
RN   [3]
RP   SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).
RC   TISSUE=Placenta;
RA   Casadei R., Strippoli P., D'Addabbo P., Canaider S., Lenzi L.,
RA   Vitale L., Giannone S., Carinci P., Zannotti M.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-130 FROM N.A.
RC   TISSUE=Smooth muscle;
RA   Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA   Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA   Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA   Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA   Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1-114 FROM N.A.
RC   TISSUE=Blood;
RA   Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA   Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT   "Human partial CDS from cd34+ stem cells.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).
RC   TISSUE=Brain;
RX   MEDLINE=99397452; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K.-I., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
RN   [7]
RP   SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).
RX   MEDLINE=92049296; PubMed=1944255;
RA   Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,
RA   Prasolov V.S.;
RT   "Identification of a protein product of a novel human gene SON and
RT   the biological effect upon administering a changed form of this gene
RT   into mammalian cells.";
RL   Mol. Biol. (Mosk) 25:731-740(1991).
RN   [8]
RP   SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).
RC   TISSUE=Placenta;
RX   MEDLINE=93062885; PubMed=1435774;
RA   Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;
RT   "The human son gene: the large and small transcripts contains various
RT   5'-terminal sequences.";
RL   Mol. Biol. (Mosk) 26:807-812(1992).
RN   [9]
RP   SEQUENCE OF 1009-1131 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=93062884; PubMed=1435773;
RA   Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,
RA   Chumakov I.M.;
RT   "Coding part of the son gene small transcript contains four areas of
RT   complete tandem repeats.";
RL   Mol. Biol. (Mosk) 26:793-806(1992).
RN   [10]
RP   SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).
RX   MEDLINE=93048367; PubMed=1424986;
RA   Mattioni T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,
RA   Lee J.S.;
RT   "A cDNA clone for a novel nuclear protein with DNA binding
RT   activity.";
RL   Chromosoma 101:618-624(1992).
RN   [11]
RP   SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).
RX   MEDLINE=89039788; PubMed=3054499;
RA   Berdichevskii F.B., Chumakov I.M., Kiselev L.L.;
RT   "Decoding of the primary structure of the son3 region in human
RT   genome: identification of a new protein with unusual structure and
RT   homology with DNA-binding proteins.";
RL   Mol. Biol. (Mosk) 22:794-801(1988).
RN   [12]
RP   SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).
RC   TISSUE=Cerebellum;
RX   MEDLINE=99439804; PubMed=10509013;
RA   Greenhalf W., Lee J., Chaudhuri B.;
RT   "A selection system for human apoptosis inhibitors using yeast.";
RL   Yeast 15:1307-1321(1999).
CC   -!- FUNCTION: Represses hepatitis B virus (HBV) core promoter activity
CC       and transcription of HBV genes and production of HBV virions.
CC       Binds to the consensus DNA sequence: 5'-GA[GT]AN[CG][AG]CC-3'.
CC       Might protect cells from apoptosis. Might be involved in pre-mRNA
CC       splicing (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear with a speckled distribution.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=10;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=F;
CC         IsoId=P18583-1; Sequence=Displayed;
CC       Name=A;
CC         IsoId=P18583-2; Sequence=VSP_004401, VSP_004402, VSP_004403;
CC       Name=B;
CC         IsoId=P18583-3; Sequence=VSP_004404, VSP_004405;
CC       Name=C;
CC         IsoId=P18583-4; Sequence=VSP_004406, VSP_004407;
CC       Name=D;
CC         IsoId=P18583-5; Sequence=VSP_004403;
CC       Name=E;
CC         IsoId=P18583-6; Sequence=VSP_004408, VSP_004409;
CC       Name=G;
CC         IsoId=P18583-7; Sequence=VSP_004410;
CC       Name=H;
CC         IsoId=P18583-8; Sequence=VSP_004411, VSP_004412;
CC       Name=I;
CC         IsoId=P18583-9; Sequence=VSP_004413;
CC       Name=J;
CC         IsoId=P18583-10; Sequence=VSP_004414, VSP_004415;
CC   -!- TISSUE SPECIFICITY: Widely expressed, with the higher expression
CC       seen in leukocyte and heart.
CC   -!- DOMAIN: Contains 8 types of repeats which are distributed in 3
CC       regions.
CC   -!- MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor
CC       SFRS2/SC-35.
CC   -!- SIMILARITY: Contains 1 G-patch domain.
CC   -!- SIMILARITY: Contains 1 DRBM (double-stranded RNA-binding) domain.
CC   -!- CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN DUE
CC       TO A FRAMESHIFT.
CC   -!- CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN
CC       DUE TO A FRAMESHIFT.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF380179; AAL34497.1; -.
DR   EMBL; X63753; CAA45282.1; ALT_FRAME.
DR   EMBL; M36428; AAA36624.1; -.
DR   EMBL; AF380180; AAL34498.1; -.
DR   EMBL; AF380181; AAL34499.1; -.
DR   EMBL; AF380182; AAL34500.1; -.
DR   EMBL; AF380183; AAL34501.1; -.
DR   EMBL; AF380184; AAL34502.1; -.
DR   EMBL; AY026895; AAK07692.1; -.
DR   EMBL; AF435977; AAL30810.1; -.
DR   EMBL; X63751; CAC69885.1; -.
DR   EMBL; AB028942; BAA82971.1; -.
DR   EMBL; X63071; CAA44793.1; ALT_FRAME.
DR   EMBL; AF139897; AAD50078.1; -.
DR   EMBL; AK024752; BAB14985.1; -.
DR   EMBL; AF161428; AAF28988.1; -.
DR   EMBL; AF161430; AAF28990.1; -.
DR   Genew; HGNC:11183; SON.
DR   GK; P18583; -.
DR   MIM; 182465; -.
DR   GO; GO:0008189; F:apoptosis inhibitor activity; IDA.
DR   GO; GO:0003677; F:DNA binding activity; TAS.
DR   GO; GO:0006916; P:anti-apoptosis; IDA.
DR   InterPro; IPR001159; DS_RBD.
DR   InterPro; IPR000467; G_patch.
DR   Pfam; PF00035; dsrm; 1.
DR   Pfam; PF01585; G-patch; 1.
DR   SMART; SM00358; DSRM; 1.
DR   SMART; SM00443; G_patch; 1.
DR   PROSITE; PS50137; DS_RBD; 1.
DR   PROSITE; PS50174; G_PATCH; 1.
KW   RNA-binding; DNA-binding; Nuclear protein; Repeat;
KW   Alternative splicing.
FT   DOMAIN      726    895       17 X 10 AA TANDEM REPEATS OF L-A-[ST]-
FT                                [NSG]-[TS]-MDSQM.

  Query Match             68.8%;  Score 75;  DB 1;  Length 2426;
  Best Local Similarity   75.0%;  Pred. No. 0.86;
  Matches   15;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              || || | |||||||:|:||
Db        109 KKHKKHKNKKKKKKKEKEKK 128


RESULT 7
Y694_METJA
ID   Y694_METJA     STANDARD;      PRT;   414 AA.
AC   Q58105;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ0694.
GN   MJ0694.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: BELONGS TO THE NOP5/NOP56 FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U67516; AAB98689.1; -.
DR   PIR; F64386; F64386.
DR   TIGR; MJ0694; -.
DR   InterPro; IPR002687; Nop.
DR   Pfam; PF01798; Nop; 1.
DR   ProDom; PD004104; Nop; 1.
KW   Hypothetical protein; Complete proteome.
FT   DOMAIN      349    414       ASP/GLU/LYS-RICH.
SQ   SEQUENCE   414 AA;  47799 MW;  A9092EFC3C82C407 CRC64;

  Query Match             67.9%;  Score 74;  DB 1;  Length 414;
  Best Local Similarity   75.0%;  Pred. No. 0.27;
  Matches   15;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              |||:|| ||:| |||| |||
Db        377 KKKEKKGKKEKSKKKKDKKK 396


RESULT 8
CBF5_KLULA
ID   CBF5_KLULA     STANDARD;      PRT;   474 AA.
AC   O13473;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE   5) (Nucleolar protein CBF5).
GN   CBF5.
OS   Kluyveromyces lactis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX   NCBI_TaxID=28985;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JBD100;
RX   MEDLINE=98144788; PubMed=9483794;
RA   Winkler A.A., Bobok A., Zonneveld B.J.M., Steensma H.Y.,
RA   Hooykaas P.J.J.;
RT   "The lysine-rich C-terminal repeats of the centromere-binding factor
RT   5 (Cbf5) of Kluyveromyces lactis are not essential for function.";
RL   Yeast 14:37-48(1998).
CC   -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC       CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC       CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC       SOME WAY ASSOCIATED WITH THE CBF3 110 kDa SUBUNIT (CBF3A) (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC   -!- SIMILARITY: Contains 1 PUA domain.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF008563; AAC64862.1; -.
DR   InterPro; IPR004802; Cbf5.
```

```
DR   InterPro; IPR002478; PUA.
DR   InterPro; IPR002501; TruB_N.
DR   InterPro; IPR004521; Unchar_dom_2.
DR   Pfam; PF01472; PUA; 1.
DR   Pfam; PF01509; TruB_N; 1.
DR   SMART; SM00359; PUA; 1.
DR   TIGRFAMs; TIGR00425; CBF5; 1.
DR   TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR   PROSITE; PS50890; PUA; 1.
KW   Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT   DOMAIN      265    340       PUA.
FT   DOMAIN      431    460       9 X 3 AA TANDEM REPEATS OF K-K-[DE].
FT   REPEAT      431    433       1.
FT   REPEAT      434    436       2.
FT   REPEAT      437    439       3.
FT   REPEAT      440    442       4.
FT   REPEAT      443    445       5.
FT   REPEAT      446    448       6.
FT   REPEAT      449    451       7.
FT   REPEAT      452    454       8.
FT   REPEAT      455    457       9.
SQ   SEQUENCE   474 AA;  53630 MW;  95306ECE7FEA756C CRC64;

  Query Match             67.9%;  Score 74;  DB 1;  Length 474;
  Best Local Similarity   70.0%;  Pred. No. 0.3;
  Matches   14;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||:||:|| ||:||:||:||
Db        434 KKEKKEKKDKKEKKEKKEKK 453


RESULT 9
CBF5_YEAST
ID   CBF5_YEAST     STANDARD;      PRT;   483 AA.
AC   P33322;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE   5) (Nucleolar protein CBF5) (P64').
GN   CBF5 OR YLR175W OR L9470.11.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=93330283; PubMed=8336724;
RA   Jiang W., Middleton K., Yoon H.-J., Fouquet C., Carbon J.;
RT   "An essential yeast protein, CBF5p, binds in vitro to centromeres and
RT   microtubules.";
RL   Mol. Cell. Biol. 13:4884-4893(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313267; PubMed=9169871;
RA   Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA   Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA   Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA   Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA   Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA   Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA   Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA   Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA   Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA   Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA   Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL   Nature 387:87-90(1997).
CC   -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC       CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC       CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC       SOME WAY ASSOCIATED WITH THE CBF3 110 kDa SUBUNIT (CBF3A).
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC   -!- SIMILARITY: Contains 1 PUA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L12351; AAA34473.1; -.
DR   EMBL; U17246; AAB67463.1; -.
DR   PIR; S41853; S41853.
DR   SGD; S0004165; CBF5.
DR   GO; GO:0005732; C:small nucleolar ribonucleoprotein complex; IPI.
DR   InterPro; IPR004802; Cbf5.
DR   InterPro; IPR002478; PUA.
DR   InterPro; IPR002501; TruB_N.
DR   InterPro; IPR004521; Unchar_dom_2.
DR   Pfam; PF01472; PUA; 1.
DR   Pfam; PF01509; TruB_N; 1.
DR   SMART; SM00359; PUA; 1.
DR   TIGRFAMs; TIGR00425; CBF5; 1.
DR   TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR   PROSITE; PS50890; PUA; 1.
KW   Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT   DOMAIN      266    341       PUA.
FT   DOMAIN      434    463       10 X 3 AA TANDEM REPEATS OF K-K-[DE].
FT   REPEAT      434    436       1.
FT   REPEAT      437    439       2.
FT   REPEAT      440    442       3.
FT   REPEAT      443    445       4.
FT   REPEAT      446    448       5.
FT   REPEAT      449    451       6.
FT   REPEAT      452    454       7.
FT   REPEAT      455    457       8.
FT   REPEAT      458    460       9.
FT   REPEAT      461    463       10.
SQ   SEQUENCE   483 AA;  54704 MW;  D356B39FDCC32E2D CRC64;

  Query Match             67.9%;  Score 74;  DB 1;  Length 483;
  Best Local Similarity   70.0%;  Pred. No. 0.31;
  Matches   14;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||:||:|| ||:||:||:||
Db        434 KKEKKEKKDKKEKKEKKEKK 453


RESULT 10
BRD3_HUMAN
ID   BRD3_HUMAN     STANDARD;      PRT;   726 AA.
AC   Q15059; Q92645;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bromodomain-containing protein 3 (RING3-like protein).
GN   BRD3 OR RING3L OR KIAA0043.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96051398; PubMed=7584044;
RA   Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA   Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. II.
RT   The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
```

```
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 1:223-229(1994).
RN   [2]
RP   SEQUENCE OF 363-726 FROM N.A.
RX   MEDLINE=98038990; PubMed=9373153;
RA   Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.;
RT   "Chromosomal localization, gene structure and transcription pattern of
RT   the ORFX gene, a homologue of the MHC-linked RING3 gene.";
RL   Gene 200:177-183(1997).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: Contains 2 bromodomains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D26362; BAA05393.1; -.
DR   EMBL; Z81330; CAB03630.1; -.
DR   HSSP; Q92831; 1B91.
DR   Genew; HGNC:1104; BRD3.
DR   MIM; 601541; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   InterPro; IPR001487; Bromodomain.
DR   Pfam; PF00439; bromodomain; 2.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 2.
DR   PROSITE; PS00633; BROMODOMAIN_1; 2.
DR   PROSITE; PS50014; BROMODOMAIN_2; 2.
KW   Bromodomain; Repeat; Nuclear protein.
FT   DOMAIN       56    115       BROMODOMAIN 1.
FT   DOMAIN      326    398       BROMODOMAIN 2.
FT   DOMAIN      487    555       LYS-RICH.
FT   DOMAIN      676    725       SER-RICH.
FT   CONFLICT    465    466       EL -> DV (IN REF. 2).
SQ   SEQUENCE   726 AA;  79541 MW;  64F526FC3C1033AA CRC64;

  Query Match             67.9%;  Score 74;  DB 1;  Length 726;
  Best Local Similarity   70.0%;  Pred. No. 0.42;
  Matches   14;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||:||:|:|||| |:|:|
Db        489 KKKKEKKEKEKKKKDKEKEK 508


RESULT 11
CBF5_CANAL
ID   CBF5_CANAL     STANDARD;      PRT;   479 AA.
AC   O43101;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Centromere/microtubule binding protein CBF5 (Centromere-binding factor
DE   5) (Nucleolar protein CBF5).
GN   CBF5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Jiang W., Clifford J., Koltin Y.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC       CENTROMERE DNA-CBF3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC       CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC       SOME WAY ASSOCIATED WITH THE CBF3 110 kDa SUBUNIT (CBF3A) (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC   -!- SIMILARITY: Contains 1 PUA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U59149; AAB94297.1; -.
DR   InterPro; IPR004802; Cbf5.
DR   InterPro; IPR002478; PUA.
DR   InterPro; IPR002501; TruB_N.
DR   InterPro; IPR004521; Unchar_dom_2.
DR   Pfam; PF01472; PUA; 1.
DR   Pfam; PF01509; TruB_N; 1.
DR   SMART; SM00359; PUA; 1.
DR   TIGRFAMs; TIGR00425; CBF5; 1.
DR   TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR   PROSITE; PS50890; PUA; 1.
KW   Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT   DOMAIN      267    342       PUA.
SQ   SEQUENCE   479 AA;  54321 MW;  3BAF5104E12C9EB6 CRC64;

  Query Match             66.1%;  Score 72;  DB 1;  Length 479;
  Best Local Similarity   70.0%;  Pred. No. 0.47;
  Matches   14;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||:|| ||:||:|| ||:||
Db        432 KKEKKDKKEKKEKKDKKEKK 451


RESULT 12
GARP_PLAFF
ID   GARP_PLAFF     STANDARD;      PRT;   678 AA.
AC   P13816;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glutamic acid-rich protein precursor.
GN   GARP.
OS   Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5837;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89040048; PubMed=2903445;
RA   Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA   Kemp D.J.;
RT   "Structure of a Plasmodium falciparum gene that encodes a glutamic
RT   acid-rich protein (GARP).";
RL   Mol. Biochem. Parasitol. 31:199-202(1988).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J03998; AAA29605.1; -.
DR   PIR; A54514; A54514.
KW   Repeat; Malaria; Antigen; Signal.
FT   SIGNAL        1     25
FT   CHAIN        26    678       GLUTAMIC ACID-RICH PROTEIN.
FT   DOMAIN      120    164       15 X 3 AA TANDEM REPEATS OF K-K-X.
FT   DOMAIN      372    416       9 X APPROXIMATE TANDEM REPEATS.
FT   DOMAIN      417    441       5 X APPROXIMATE TANDEM REPEATS.
```

```
FT   DOMAIN      576    604       POLY-GLU.
FT   DOMAIN      605    653       7 X APPROXIMATE TANDEM REPEATS.
FT   DOMAIN      654    663       POLY-GLU.
SQ   SEQUENCE   678 AA;  80551 MW;  2A8F85606496EA9E CRC64;

  Query Match             66.1%;  Score 72;  DB 1;  Length 678;
  Best Local Similarity   70.0%;  Pred. No. 0.62;
  Matches   14;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||:||:|| ||:|| ||:||
Db        132 KKEKKEKKDKKEKKDKKEKK 151


RESULT 13
CG79_HUMAN
ID   CG79_HUMAN     STANDARD;      PRT;   351 AA.
AC   Q9Y388;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein CGI-79.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20272150; PubMed=10810093;
RA   Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT   "Identification of novel human genes evolutionarily conserved in
RT   Caenorhabditis elegans by comparative proteomics.";
RL   Genome Res. 10:703-713(2000).
CC   -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF151837; AAD34074.1; -.
DR   HSSP; P26368; 2U2F.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Hypothetical protein; RNA-binding.
FT   DOMAIN       36    114       RNA-BINDING (RRM).
SQ   SEQUENCE   351 AA;  39675 MW;  7B6E882D6B192EBE CRC64;

  Query Match             65.1%;  Score 71;  DB 1;  Length 351;
  Best Local Similarity   73.7%;  Pred. No. 0.46;
  Matches   14;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKK 20
              || || ||:||||||:|:|
Db        156 KKHKKDKKEKKKKKKEKEK 174


RESULT 14
CNG1_HUMAN
ID   CNG1_HUMAN     STANDARD;      PRT;   686 AA.
AC   P29973; Q16279; Q16485;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   cGMP-gated cation channel alpha 1 (CNG channel alpha 1) (CNG-1) (CNG1)
DE   (Cyclic nucleotide gated channel alpha 1) (Cyclic nucleotide gated
DE   channel, photoreceptor) (Cyclic-nucleotide-gated cation channel 1)
DE   (Rod photoreceptor cGMP-gated channel alpha subunit).
GN   CNGA1 OR CNCG1 OR CNCG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=92210603; PubMed=1372902;
RA   Pittler S.J., Lee A.K., Altherr M.R., Howard T.A., Seldin M.F.,
RA   Hurwitz R.L., Wasmuth J.J., Baehr W.;
RT   "Primary structure and chromosomal localization of human and mouse
RT   rod photoreceptor cGMP-gated cation channel.";
RL   J. Biol. Chem. 267:6257-6262(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=92356211; PubMed=1379636;
RA   Dhallan R.S., Macke J.P., Eddy R.L., Shows T.B., Reed R.R.,
RA   Yau K.-W., Nathans J.;
RT   "Human rod photoreceptor cGMP-gated channel: amino acid sequence,
RT   gene structure, and functional expression.";
RL   J. Neurosci. 12:3248-3256(1992).
RN   [3]
RP   SEQUENCE OF 313-573 FROM N.A.
RX   MEDLINE=95175019; PubMed=7532814;
RA   Distler M., Biel M., Flockerzi V., Hofmann F.;
RT   "Expression of cyclic nucleotide-gated cation channels in non-sensory
RT   tissues and cells.";
RL   Neuropharmacology 33:1275-1282(1994).
RN   [4]
RP   VARIANT ARRP PHE-316, AND VARIANTS GLN-28 AND ASN-114.
RX   MEDLINE=96036047; PubMed=7479749;
RA   Dryja T.P., Finn J.T., Peng Y.-W., McGee T.L., Berson E.L., Yau K.-W.;
RT   "Mutations in the gene encoding the alpha subunit of the rod
RT   cGMP-gated channel in autosomal recessive retinitis pigmentosa.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:10177-10181(1995).
CC   -!- FUNCTION: VISUAL SIGNAL TRANSDUCTION IS MEDIATED BY A G-PROTEIN
CC       COUPLED CASCADE USING CGMP AS SECOND MESSENGER. THIS PROTEIN CAN
CC       BE ACTIVATED BY CYCLIC GMP WHICH LEADS TO AN OPENING OF THE CATION
CC       CHANNEL AND THEREBY CAUSING A DEPOLARIZATION OF ROD
CC       PHOTORECEPTORS.
CC   -!- SUBUNIT: HOMOTETRAMER OR HIGHER OLIGOMER. FORMS HETEROOLIGOMERIC
CC       COMPLEX WITH CNG4.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: ROD CELLS IN THE RETINA.
CC   -!- DISEASE: Defects in CNGA1 are a cause of autosomal recessive
CC       retinitis pigmentosa (ARRP) [MIM:123825], a disease that leads to
CC       degeneration of retinal photoreceptor cells.
CC   -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL
CC       FAMILY.
CC   -!- DATABASE: NAME=Mutations of the CNGA1 gene;
CC       NOTE=Retina International's Scientific Newsletter;
CC       WWW="http://www.retina-international.com/sci-news/cnga1mut.htm".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M84741; AAA52010.1; ALT_INIT.
DR   EMBL; S42457; AAB22778.1; -.
DR   EMBL; S76062; AAD14206.1; -.
DR   Genew; HGNC:2148; CNGA1.
DR   MIM; 123825; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0006832; P:small molecule transport; TAS.
DR   GO; GO:0007601; P:vision; TAS.
DR   InterPro; IPR000595; cNMP_binding.
```

```
DR   InterPro; IPR005821; Ion_trans.
DR   InterPro; IPR001622; K+channel_pore.
DR   Pfam; PF00027; cNMP_binding; 1.
DR   Pfam; PF00520; ion_trans; 1.
DR   SMART; SM00100; cNMP; 1.
DR   PROSITE; PS00888; CNMP_BINDING_1; 1.
DR   PROSITE; PS00889; CNMP_BINDING_2; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
KW   Ionic channel; Ion transport; cAMP-binding; Transmembrane;
KW   Multigene family; Vision; Disease mutation; Polymorphism;
KW   Retinitis pigmentosa.
FT   DOMAIN        1    160       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    161    181       H1 (POTENTIAL).
FT   DOMAIN      182    194       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    195    213       H2 (POTENTIAL).
FT   DOMAIN      214    237       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    238    257       H3 (POTENTIAL).
FT   DOMAIN      258    295       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    296    318       H4 (POTENTIAL).
FT   DOMAIN      319    370       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    371    390       H5 (POTENTIAL).
FT   DOMAIN      391    474       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    475    495       H6 (POTENTIAL).
FT   DOMAIN      496    686       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     483    605       CGMP (POTENTIAL).
FT   BINDING     542    542       CGMP (POTENTIAL).
FT   BINDING     557    557       CGMP (POTENTIAL).
FT   CARBOHYD    421    421       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT      28     28       R -> Q.
FT                                /FTId=VAR_009295.
FT   VARIANT     114    114       D -> N.
FT                                /FTId=VAR_009296.
FT   VARIANT     316    316       S -> F (in ARRP).
FT                                /FTId=VAR_009297.
FT   CONFLICT     46     46       S -> Y (IN REF. 1).
FT   CONFLICT     85     85       L -> I (IN REF. 1).
FT   CONFLICT    146    147       EE -> HH (IN REF. 1).
FT   CONFLICT    539    539       Y -> T (IN REF. 1).
FT   CONFLICT    677    678       GA -> WS (IN REF. 1).
SQ   SEQUENCE   686 AA;  79126 MW;  E5200D216FC97AF6 CRC64;

  Query Match             65.1%;  Score 71;  DB 1;  Length 686;
  Best Local Similarity   70.0%;  Pred. No. 0.78;
  Matches   14;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||  |:||||::| | |
Db        124 KKKKKKDKEKKKKEEKSKDK 143


RESULT 15
TCOF_HUMAN
ID   TCOF_HUMAN     STANDARD;      PRT;  1411 AA.
AC   Q13428; Q99408; Q99860;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Treacle protein (Treacher Collins syndrome protein).
GN   TCOF1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96154183; PubMed=8563749;
RA   Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,
RA   Bonner C.A., Koprivnikar K., Wasmuth J.J.;
RT   "Positional cloning of a gene involved in the pathogenesis of
RT   Treacher Collins syndrome.";
RL   Nat. Genet. 12:130-136(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97228900; PubMed=9074926;
RA   Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,
RA   Dixon M.J.;
RT   "Identification of the complete coding sequence and genomic
RT   organization of the Treacher Collins syndrome gene.";
RL   Genome Res. 7:223-234(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97250498; PubMed=9096354;
RA   Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
RA   Ashley J.A., Lovett M., Jabs E.W.;
RT   "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits
RT   mutations in Treacher Collins syndrome throughout its coding
RT   region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
RN   [4]
RP   VARIANTS LEU-439; VAL-810; VAL-1313 AND GLY-1355, AND VARIANT TCS
RP   ARG-53.
RX   MEDLINE=97195537; PubMed=9042910;
RA   Edwards S.J., Gladwin A.J., Dixon M.J.;
RT   "The mutational spectrum in Treacher Collins syndrome reveals a
RT   predominance of mutations that create a premature-termination
RT   codon.";
RL   Am. J. Hum. Genet. 60:515-524(1997).
CC   -!- DISEASE: Defects in TCOF1 are the cause of Treacher Collins
CC       syndrome (TCS) [MIM:154500]. TCS is an autosomal dominant disorder
CC       of craniofacial development that occurs with an incidence of
CC       1/50,000 live births. The clinical features of TCS are bilaterally
CC       symmetrical and include: (1) abnormalities of the external ears,
CC       atresia of the external ear canals, and malformation of the middle
CC       ear ossicles, which may result in conductive hearing loss; (2)
CC       lateral downward sloping of palpebral fissures, frequently with
CC       colobomas of the lower eyelids; (3) hypoplasia of the mandible and
CC       zygomatic complex; (4) cleft palate.
CC   -!- SIMILARITY: Contains 1 LisH domain.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U40847; AAC50903.1; -.
DR   EMBL; U76366; AAC51181.1; -.
DR   EMBL; U84664; AAC51185.1; -.
DR   EMBL; U84640; AAC51185.1; JOINED.
DR   EMBL; U84641; AAC51185.1; JOINED.
DR   EMBL; U84642; AAC51185.1; JOINED.
DR   EMBL; U84643; AAC51185.1; JOINED.
DR   EMBL; U84644; AAC51185.1; JOINED.
DR   EMBL; U84645; AAC51185.1; JOINED.
DR   EMBL; U84646; AAC51185.1; JOINED.
DR   EMBL; U84647; AAC51185.1; JOINED.
DR   EMBL; U84648; AAC51185.1; JOINED.
DR   EMBL; U84649; AAC51185.1; JOINED.
DR   EMBL; U84650; AAC51185.1; JOINED.
DR   EMBL; U84651; AAC51185.1; JOINED.
DR   EMBL; U84652; AAC51185.1; JOINED.
DR   EMBL; U84653; AAC51185.1; JOINED.
DR   EMBL; U84654; AAC51185.1; JOINED.
DR   EMBL; U84655; AAC51185.1; JOINED.
DR   EMBL; U84656; AAC51185.1; JOINED.
DR   EMBL; U84657; AAC51185.1; JOINED.
DR   EMBL; U84658; AAC51185.1; JOINED.
DR   EMBL; U84659; AAC51185.1; JOINED.
DR   EMBL; U84660; AAC51185.1; JOINED.
DR   EMBL; U84661; AAC51185.1; JOINED.
DR   EMBL; U84662; AAC51185.1; JOINED.
DR   EMBL; U84663; AAC51185.1; JOINED.
DR   EMBL; U79659; AAB40722.1; -.
DR   EMBL; U79645; AAB40722.1; JOINED.
```

```
DR   EMBL; U79646; AAB40722.1; JOINED.
DR   EMBL; U79647; AAB40722.1; JOINED.
DR   EMBL; U79648; AAB40722.1; JOINED.
DR   EMBL; U79649; AAB40722.1; JOINED.
DR   EMBL; U79650; AAB40722.1; JOINED.
DR   EMBL; U79651; AAB40722.1; JOINED.
DR   EMBL; U79652; AAB40722.1; JOINED.
DR   EMBL; U79653; AAB40722.1; JOINED.
DR   EMBL; U79654; AAB40722.1; JOINED.
DR   EMBL; U79655; AAB40722.1; JOINED.
DR   EMBL; U79656; AAB40722.1; JOINED.
DR   EMBL; U79657; AAB40722.1; JOINED.
DR   EMBL; U79658; AAB40722.1; JOINED.
DR   Genew; HGNC:11654; TCOF1.
DR   MIM; 606847; -.
DR   MIM; 154500; -.
DR   GO; GO:0005730; C:nucleolus; TAS.
DR   GO; GO:0005215; F:transporter activity; TAS.
DR   GO; GO:0001501; P:skeletal development; TAS.
DR   InterPro; IPR006594; LisH.
DR   InterPro; IPR003993; treacle.
DR   Pfam; PF03546; treacle; 3.
DR   PRINTS; PR01503; TREACLE.
DR   SMART; SM00667; LisH; 1.
DR   PROSITE; PS50896; LISH; 1.
KW   Disease mutation; Polymorphism.
FT   DOMAIN        6     38       LISH.
FT   DOMAIN       89     97       POLY-GLU.
FT   DOMAIN      204    207       POLY-SER.
FT   DOMAIN      616    619       POLY-SER.
FT   DOMAIN      919    924       POLY-SER.
FT   DOMAIN     1285   1289       POLY-LYS.
FT   DOMAIN     1375   1386       POLY-LYS.
FT   DOMAIN     1398   1405       POLY-LYS.
FT   VARIANT      53     53       W -> R (in TCS).
FT                                /FTId=VAR_005630.
FT   VARIANT     439    439       P -> L.
FT                                /FTId=VAR_005631.
FT   VARIANT     810    810       A -> V.
FT                                /FTId=VAR_005632.
FT   VARIANT    1313   1313       A -> V (in dbSNP:15251).
FT                                /FTId=VAR_005633.
FT   VARIANT    1355   1355       D -> G.
FT                                /FTId=VAR_005634.
FT   CONFLICT   1312   1312       K -> Q (IN REF. 2).
SQ   SEQUENCE   1411 AA;  144312 MW;  3880203D985C2699 CRC64;

  Query Match             64.7%;  Score 70.5;  DB 1;  Length 1411;
  Best Local Similarity   54.8%;  Pred. No. 1.5;
  Matches   17;  Conservative    2;  Mismatches    1;  Indels   11;  Gaps    1;

Qy          2 KKKKKKKKKKKK-----------KKKKKKKK 21
              |:||:|||| ||            ||||||||
Db       1375 KEKKEKKKKAKKASTKDSESPSQKKKKKKKK 1405


Search completed: January 30, 2004, 00:20:44
Job time : 5.73239 secs
```

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:58:52 ; Search time 21.6901 Seconds
(without alignments)
249.842 Million cell updates/sec

Title: US-09-461-684C-1
Perfect score: 109
Sequence: 1 CKKKKKKKKKKKKKKKKKKKK 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	100	91.7	55	4	Q8N6F0	Q8n6f0 homo sapien
2	100	91.7	80	10	Q8S7D3	Q8s7d3 oryza sativ
3	100	91.7	113	10	Q8LQP6	Q8lqp6 oryza sativ
4	100	91.7	128	3	Q9P529	Q9p529 neurospora
5	100	91.7	129	11	O35807	O35807 rattus norv
6	100	91.7	168	4	Q9H5V6	Q9h5v6 homo sapien
7	100	91.7	206	5	Q8I247	Q8i247 plasmodium
8	100	91.7	215	11	Q64075	Q64075 rattus sp.
9	100	91.7	260	10	Q9LGZ9	Q9lgz9 arabidopsis
10	100	91.7	380	4	Q9NT34	Q9nt34 homo sapien
11	100	91.7	515	5	Q8SWR7	Q8swr7 drosophila
12	100	91.7	517	10	Q9LXR2	Q9lxr2 arabidopsis
13	100	91.7	531	6	Q95LV6	Q95lv6 macaca fasc
14	100	91.7	720	4	Q9H6Q7	Q9h6q7 homo sapien
15	100	91.7	791	5	Q8T2U7	Q8t2u7 dictyosteli
16	97	89.0	667	4	Q9HC48	Q9hc48 homo sapien
17	96	88.1	476	10	Q8RVV4	Q8rvv4 lycopersico
18	95	87.2	489	5	Q8T3H4	Q8t3h4 drosophila
19	95	87.2	658	11	Q8CGI8	Q8cgi8 mus musculu
20	92	84.4	4433	5	Q8IJI5	Q8iji5 plasmodium
21	91	83.5	3351	5	Q8IBK4	Q8ibk4 plasmodium
22	90	82.6	39	5	Q8MSJ8	Q8msj8 drosophila
23	90	82.6	257	4	Q9H5Y3	Q9h5y3 homo sapien
24	90	82.6	686	4	Q9NXF0	Q9nxf0 homo sapien
25	90	82.6	2820	5	Q8IM32	Q8im32 plasmodium
26	89	81.7	83	11	Q9ER82	Q9er82 mus musculu
27	89	81.7	1710	5	Q8I239	Q8i239 plasmodium
28	88	80.7	109	5	Q8I3G8	Q8i3g8 plasmodium
29	87	79.8	107	10	Q9LQF6	Q9lqf6 arabidopsis
30	87	79.8	476	5	Q8IFN3	Q8ifn3 plasmodium
31	86	78.9	102	11	Q8C5Y8	Q8c5y8 mus musculu
32	86	78.9	741	5	Q8IPR6	Q8ipr6 drosophila
33	85	78.0	978	5	Q8IB83	Q8ib83 plasmodium
34	84	77.1	339	4	Q8IVS7	Q8ivs7 homo sapien
35	83	76.1	227	4	Q8TA86	Q8ta86 homo sapien
36	83	76.1	1337	5	Q8IC23	Q8ic23 plasmodium
37	82.5	75.7	710	5	Q8IDM7	Q8idm7 plasmodium
38	82	75.2	943	5	Q8ILY0	Q8ily0 plasmodium
39	81	74.3	253	11	Q8CFH9	Q8cfh9 mus musculu
40	81	74.3	547	4	Q8IV81	Q8iv81 homo sapien
41	81	74.3	1082	5	Q8I5Y6	Q8i5y6 plasmodium
42	81	74.3	1313	5	Q8IKC0	Q8ikc0 plasmodium
43	81	74.3	3193	5	Q8I590	Q8i590 plasmodium
44	81	74.3	4550	5	O77336	O77336 plasmodium
45	80	73.4	213	11	P97762	P97762 mus musculu

ALIGNMENTS

RESULT 1
Q8N6F0
ID Q8N6F0 PRELIMINARY; PRT; 55 AA.
AC Q8N6F0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to LOC201361.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC030525; AAH30525.1; -.
SQ SEQUENCE 55 AA; 7251 MW; 0906032B284006BA CRC64;

Query Match 91.7%; Score 100; DB 4; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.9e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db         21 KKKKKKKKKKKKKKKKKKKK 40
```

RESULT 2
Q8S7D3
ID Q8S7D3 PRELIMINARY; PRT; 80 AA.
AC Q8S7D3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 9.4 kDa protein.
GN OSJNBA0057L21.23.
OS Oryza sativa (Rice).

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA   Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA   Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA   VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA   Salzberg S.L., White O., Fraser C.M.;
RT   "Oryza sativa chromosome 10 BAC OSJNBa0057L21 genomic sequence.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC087599; AAL79706.1; -.
DR   Gramene; Q8S7D3; -.
KW   Hypothetical protein.
SQ   SEQUENCE   80 AA;  9362 MW;  0177C863133B21D8 CRC64;

  Query Match             91.7%;  Score 100;  DB 10;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 0.0001;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db         48 KKKKKKKKKKKKKKKKKKKK 67

RESULT 3
Q8LQP6
ID   Q8LQP6      PRELIMINARY;      PRT;   113 AA.
AC   Q8LQP6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   OJ1117_G01.13 protein.
GN   OJ1117_G01.13.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:OJ1117_G01.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003374; BAB93330.1; -.
DR   Gramene; Q8LQP6; -.
SQ   SEQUENCE   113 AA;  13660 MW;  597DB0EDEB2AA3EF CRC64;

  Query Match             91.7%;  Score 100;  DB 10;  Length 113;
  Best Local Similarity   100.0%;  Pred. No. 0.00013;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db         10 KKKKKKKKKKKKKKKKKKKK 29

RESULT 4
Q9P529
ID   Q9P529      PRELIMINARY;      PRT;   128 AA.
AC   Q9P529;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 15.2 kDa protein.
GN   B24H17.160.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356815; CAB92638.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   128 AA;  15157 MW;  8C7C65C3DFB70765 CRC64;

  Query Match             91.7%;  Score 100;  DB 3;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 0.00014;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db         71 KKKKKKKKKKKKKKKKKKKK 90

RESULT 5
O35807
ID   O35807      PRELIMINARY;      PRT;   129 AA.
AC   O35807;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   MICROVASCULAR endothelial differentiation protein 2.
GN   MDG2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Epididymis;
RX   MEDLINE=98172708; PubMed=9511718;
RA   Proels F., Loser B., Marx M.;
RT   "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA
RT   species, during in vitro angiogenesis.";
RL   Exp. Cell Res. 239:1-10(1998).
DR   EMBL; Y08769; CAA70022.1; -.
DR   InterPro; IPR000719; Prot_kinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   129 AA;  15080 MW;  38102272BBE2EDB4 CRC64;

  Query Match             91.7%;  Score 100;  DB 11;  Length 129;
  Best Local Similarity   100.0%;  Pred. No. 0.00014;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db         85 KKKKKKKKKKKKKKKKKKKK 104

RESULT 6
Q9H5V6
ID   Q9H5V6      PRELIMINARY;      PRT;   168 AA.
AC   Q9H5V6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ22976 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA   Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA   Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK026629; BAB15513.1; -.
KW   Hypothetical protein.
FT   NON_TER     168    168
SQ   SEQUENCE   168 AA;  19549 MW;  A19DBD195F8A1A90 CRC64;

  Query Match             91.7%;  Score 100;  DB 4;  Length 168;
  Best Local Similarity   100.0%;  Pred. No. 0.00016;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        140 KKKKKKKKKKKKKKKKKKKK 159


RESULT 7
Q8I247
ID   Q8I247      PRELIMINARY;      PRT;   206 AA.
AC   Q8I247;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFA0475C.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22255708; PubMed=12368867;
RA   Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA   Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA   Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA   Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA   Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA   Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA   Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA   Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA   Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA   Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA   Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,
RA   Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA   Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA   Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA   Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT   "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL   Nature 419:527-531(2002).
DR   EMBL; AL031745; CAD49055.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   206 AA;  25047 MW;  1192E49A3DC4523F CRC64;

  Query Match             91.7%;  Score 100;  DB 5;  Length 206;
  Best Local Similarity   100.0%;  Pred. No. 0.00019;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        185 KKKKKKKKKKKKKKKKKKKK 204


RESULT 8
Q64075
ID   Q64075      PRELIMINARY;      PRT;   215 AA.
AC   Q64075;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Nucleoporin p62 homolog protein (Fragment).
OS   Rattus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10118;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95151924; PubMed=7849178;
RA   Wang Z.Q., Akmal K.M., Kim K.H.;
RT   "An unusual nucleoporin-related messenger ribonucleic acid is present
RT   in the germ cells of rat testis.";
RL   Biol. Reprod. 51:1022-1030(1994).
DR   EMBL; S75997; AAB33384.1; -.
KW   Porin.
FT   NON_TER       1      1
SQ   SEQUENCE   215 AA;  24593 MW;  098251C97A8FBD88 CRC64;

  Query Match             91.7%;  Score 100;  DB 11;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 0.00019;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db         35 KKKKKKKKKKKKKKKKKKKK 54


RESULT 9
Q9LGZ9
ID   Q9LGZ9      PRELIMINARY;      PRT;   260 AA.
AC   Q9LGZ9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Genomic DNA, chromosome 3, BAC clone:F1D9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RA   Nakamura Y.;
RT   "Structural Analysis of Arabidopsis thaliana Chromosome 3. III.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002460; BAA97098.1; -.
DR   InterPro; IPR005819; Histone_H5.
DR   PRINTS; PR00624; HISTONEH5.
SQ   SEQUENCE   260 AA;  33307 MW;  43E2394CB8131143 CRC64;

  Query Match             91.7%;  Score 100;  DB 10;  Length 260;
  Best Local Similarity   100.0%;  Pred. No. 0.00022;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db          7 KKKKKKKKKKKKKKKKKKKK 26


RESULT 10
Q9NT34
ID   Q9NT34      PRELIMINARY;      PRT;   380 AA.
AC   Q9NT34;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP434I1120.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137556; CAB70810.1; -.
DR   Genew; HGNC:15736; C17orf28.
KW   Hypothetical protein.
FT   NON_TER     380    380
SQ   SEQUENCE   380 AA;  42689 MW;  67F50DD101346AFB CRC64;

  Query Match             91.7%;  Score 100;  DB 4;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 0.00028;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        355 KKKKKKKKKKKKKKKKKKKK 374


RESULT 11
Q8SWR7
ID   Q8SWR7      PRELIMINARY;      PRT;   515 AA.
AC   Q8SWR7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   GH22607p (Fragment).
GN   CG7180.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY095518; AAM12251.1; -.
DR   FlyBase; FBgn0032673; CG7180.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_PP.
DR   Pfam; PF00102; Y_phosphatase; 1.
DR   SMART; SM00194; PTPc; 1.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW   Hydrolase.
FT   NON_TER     515    515
SQ   SEQUENCE   515 AA;  59080 MW;  B2825B7EEA96195E CRC64;

  Query Match             91.7%;  Score 100;  DB 5;  Length 515;
  Best Local Similarity   100.0%;  Pred. No. 0.00034;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        493 KKKKKKKKKKKKKKKKKKKK 512


RESULT 12
Q9LXR2
ID   Q9LXR2      PRELIMINARY;      PRT;   517 AA.
AC   Q9LXR2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 59.7 kDa protein.
GN   T20N10_250.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA   Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353032; CAB88307.1; -.
DR   InterPro; IPR001810; F-box.
DR   InterPro; IPR006566; FBD.
DR   Pfam; PF00646; F-box; 1.
DR   SMART; SM00579; FBD; 1.
DR   SMART; SM00256; FBOX; 1.
DR   PROSITE; PS50181; FBOX; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   517 AA;  59689 MW;  EC6D957D01F86E70 CRC64;

  Query Match             91.7%;  Score 100;  DB 10;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 0.00034;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        444 KKKKKKKKKKKKKKKKKKKK 463


RESULT 13
Q95LV6
ID   Q95LV6      PRELIMINARY;      PRT;   531 AA.
AC   Q95LV6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 61.4 kDa protein (Fragment).
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA   Terao K., Sugano S.;
RT   "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT   libraries.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB071085; BAB64479.1; -.
KW   Hypothetical protein.
FT   NON_TER     531    531
SQ   SEQUENCE   531 AA;  61389 MW;  B55996B4F5CDD60C CRC64;

  Query Match             91.7%;  Score 100;  DB 6;  Length 531;
  Best Local Similarity   100.0%;  Pred. No. 0.00035;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        502 KKKKKKKKKKKKKKKKKKKK 521


RESULT 14
Q9H6Q7
ID   Q9H6Q7      PRELIMINARY;      PRT;   720 AA.
```

```
AC   Q9H6Q7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ21979 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK025632; BAB15196.1; -.
KW   Hypothetical protein.
FT   NON_TER     720    720
SQ   SEQUENCE   720 AA;  84029 MW;  A86586FEAA953D0B CRC64;

  Query Match             91.7%;  Score 100;  DB 4;  Length 720;
  Best Local Similarity   100.0%;  Pred. No. 0.00043;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        692 KKKKKKKKKKKKKKKKKKKK 711

RESULT 15
Q8T2U7
ID   Q8T2U7      PRELIMINARY;      PRT;   791 AA.
AC   Q8T2U7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 92.4 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115574; AAL92183.1; -.
DR   InterPro; IPR005033; YEATS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF03366; YEATS; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   791 AA;  92375 MW;  D66CCB6DEC92352C CRC64;

  Query Match             91.7%;  Score 100;  DB 5;  Length 791;
  Best Local Similarity   100.0%;  Pred. No. 0.00045;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 KKKKKKKKKKKKKKKKKKKK 21
              ||||||||||||||||||||
Db        769 KKKKKKKKKKKKKKKKKKKK 788

Search completed: January 30, 2004, 00:24:37
Job time : 22.6901 secs
```

OM protein - protein search, using sw model

```
Run on:         January 30, 2004, 00:13:12 ; Search time 10.6808 Seconds
                                             (without alignments)
                                             225.098 Million cell updates/sec

Title:          US-09-461-684C-2
Perfect score:  109
Sequence:       1 CEAAAAAEAAAAAEAAAAAEAAAAA 25

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_76:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	77	70.6	265	2	S19113	cgcr-4 protein - C
2	72	66.1	165	2	B87702	ribosomal protein
3	71	65.1	205	2	S19114	cgcr-1 protein - C
4	70	64.2	183	2	S24960	gene C98 protein -
5	70	64.2	401	2	A48423	engrailed homeodom
6	69	63.3	97	2	S02376	antifreeze protein
7	69	63.3	1028	2	A56038	DNA-binding protei
8	69	63.3	1213	2	S16356	ovo protein - frui
9	69	63.3	1668	2	T13748	sex comb protein -
10	68	62.4	403	2	A81882	probable dihydroli
11	68	62.4	655	2	A29945	neurogenesis regul
12	68	62.4	873	2	B53225	ecdysone-induced p
13	68	62.4	2023	2	T13154	polycomb protein e
14	67	61.5	314	2	JC5273	paired type homeob
15	67	61.5	460	2	T33110	hypothetical prote
16	67	61.5	494	2	A42170	zinc finger protei
17	67	61.5	497	2	JC5076	myc-associated zin
18	67	61.5	604	2	A39369	homeotic protein B
19	67	61.5	606	2	S13367	Om(1D) protein - f
20	67	61.5	627	2	T02610	probable YME1 ATP-
21	67	61.5	671	2	C96534	probable Poly-A Bi
22	67	61.5	1226	2	T24045	hypothetical prote
23	67	61.5	2715	2	T13049	eyelid - fruit fly
24	66.5	61.0	543	2	B39369	homeotic protein B
25	65.5	60.1	273	2	T51010	hypothetical prote
26	65	59.6	85	1	FDFL4W	antifreeze protein
27	65	59.6	91	2	A22592	antifreeze protein
28	65	59.6	333	2	A39065	homeotic protein E
29	65	59.6	475	2	A43915	homeotic protein e
30	65	59.6	644	2	S39356	transcription fact
31	65	59.6	703	2	T48600	kinase-like protei
32	64	58.7	109	1	R6UTP1	acidic ribosomal p
33	64	58.7	392	2	B48423	homeotic protein e
34	64	58.7	1065	2	T13230	dachshund isoform
35	64	58.7	1072	2	T13232	dachshund protein
36	64	58.7	1074	2	T13229	dachshund protein
37	64	58.7	1081	2	T13231	dachshund protein
38	64	58.7	1533	2	A46221	abdominal segment
39	64	58.7	2038	2	A43742	female sterile hom
40	63	57.8	179	2	AF2908	50S ribosomal prot
41	63	57.8	179	2	F97683	50S ribosomal prot
42	63	57.8	513	2	A48233	polyomavirus enhan
43	63	57.8	568	2	T39675	asparaginyl-trna s
44	63	57.8	581	2	E75383	conserved hypothet
45	63	57.8	846	2	S52418	GTP-binding regula

ALIGNMENTS

```
RESULT 1
S19113
cgcr-4 protein - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C;Accession: S19113; S14466
R;Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A;Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A;Reference number: S19113; MUID:92119224; PMID:1731966
A;Accession: S19113
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <WAK>
A;Cross-references: EMBL:X17208; NID:g18136; PIDN:CAA35080.1; PID:g18137
C;Genetics:
A;Gene: cgcr-4

  Query Match             70.6%;  Score 77;  DB 2;  Length 265;
  Best Local Similarity   82.6%;  Pred. No. 0.25;
  Matches   19;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              |||||||||||:|  |||| |||
Db        154 AAAAAEAAAAAKARVAAEARAAA 176


RESULT 2
B87702
ribosomal protein S16 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: B87702
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolona
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87702
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-165 <STO>
A;Cross-references: GB:AE005673; NID:g13425408; PIDN:AAK25614.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC3652

  Query Match             66.1%;  Score 72;  DB 2;  Length 165;
  Best Local Similarity   75.0%;  Pred. No. 0.56;
  Matches   18;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
```

```
              :|  | |:|||  ||  |||||||||
Db        115 QAEADAKAAAEAEKAAAAEAAAAA 138

RESULT 3
S19114
cgcr-1 protein - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 03-Nov-2000
C;Accession: S19114
R;Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A;Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A;Reference number: S19113; MUID:92119224; PMID:1731966
A;Accession: S19114
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-205 <WAK>
A;Cross-references: EMBL:X17207
C;Superfamily: phage lambda hypothetical protein 401

  Query Match             65.1%;  Score 71;  DB 2;  Length 205;
  Best Local Similarity   75.0%;  Pred. No. 0.82;
  Matches   18;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              | ||||:|||||||||    || |
Db         49 EEAAAAQAAAAAEAAAAEARAAPA 72

RESULT 4
S24960
gene C98 protein - rape
C;Species: Brassica napus (rape)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C;Accession: S24960
R;Roberts, M.R.; Hodge, R.; Sorensen, A.; Ross, J.; Murphy, D.J.; Draper, R.; Scott, R.
submitted to the EMBL Data Library, July 1992
A;Description: A new class of Brassica napus oleosin genes specific to the male gametoph
A;Reference number: S24960
A;Accession: S24960
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-183 <ROB>
A;Cross-references: EMBL:X67142; NID:g17792; PID:g17793

  Query Match             64.2%;  Score 70;  DB 2;  Length 183;
  Best Local Similarity   78.3%;  Pred. No. 0.95;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              || ||| | ||||| ||||| ||
Db        151 AAPAAEPAPAAEAAPAAEAAPAA 173

RESULT 5
A48423
engrailed homeodomain-containing protein En-1 - mouse
N;Alternate names: homeotic protein En-1
C;Species: Mus musculus (house mouse)
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A48423; S13009; A26629; A24778
R;Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallainathan, D.; Provart, N.J.; Joyner, A.L
Dev. Genet. 13, 345-358, 1992
A;Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene
A;Reference number: A48423; MUID:93185339; PMID:1363401
A;Accession: A48423
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-401 <LOG>
A;Experimental source: CD-1, embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:126620)
R;Holland, P.W.H.; Williams, N.A.
```

```
FEBS Lett. 277, 250-252, 1990
A;Title: Conservation of engrailed-like homeobox sequences during vertebrate evolution.
A;Reference number: S13009; MUID:91099509; PMID:1980115
A;Accession: S13009
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 321-380 <HOL>
R;Joyner, A.L.; Martin, G.R.
Genes Dev. 1, 29-38, 1987
A;Title: En-1 and En-2, two mouse genes with sequence homology to the Drosophila engraile
A;Reference number: A91620; MUID:88112776; PMID:2892757
A;Accession: A26629
A;Molecule type: DNA; mRNA
A;Residues: 278-401 <JOY>
A;Cross-references: GB:Y00201; GB:M11987; NID:g49587; PIDN:CAA68361.1; PID:g669105
R;Joyner, A.L.; Kornberg, T.; Coleman, K.G.; Cox, D.R.; Martin, G.R.
Cell 43, 29-37, 1985
A;Title: Expression during embryogenesis of a mouse gene with sequence homology to the Dr
A;Reference number: A24778; MUID:86079501; PMID:2416459
A;Accession: A24778
A;Molecule type: DNA
A;Residues: 311-401 <JO2>
C;Genetics:
A;Gene: en.1
A;Map position: 1
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;313-369/Domain: homeobox homology <HOX>

  Query Match             64.2%;  Score 70;  DB 2;  Length 401;
  Best Local Similarity   78.3%;  Pred. No. 1.7;
  Matches   18;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              ||||| ||||| || || ||||:
Db        207 AAAAAAAAAAAAAAVAAAAAAAS 229

RESULT 6
S02376
antifreeze protein precursor - yellowtail flounder
C;Species: Limanda ferruginea (yellowtail flounder)
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 24-Oct-2000
C;Accession: S02376
R;Scott, G.K.; Davies, P.L.; Shears, M.A.; Fletcher, G.L.
Eur. J. Biochem. 168, 629-633, 1987
A;Title: Structural variations in the alanine-rich antifreeze proteins of the Pleuronecti
A;Reference number: S02376; MUID:88029483; PMID:3665937
A;Accession: S02376
A;Molecule type: mRNA
A;Residues: 1-97 <SCO>
A;Cross-references: EMBL:X06356; NID:g64041; PIDN:CAA29655.1; PID:g64042
A;Note: part of this sequence, including the amino end of the mature protein, was confirm
C;Superfamily: antifreeze protein
C;Keywords: antifreeze
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-48/Domain: propeptide #status predicted <PRO>
F;49-96/Product: antifreeze protein #status predicted <MAT>

  Query Match             63.3%;  Score 69;  DB 2;  Length 97;
  Best Local Similarity   70.8%;  Pred. No. 0.75;
  Matches   17;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              :||||| | ||| | |||: ||||
Db         53 DAAAAAAATAAAAAKAAADTAAAA 76

RESULT 7
A56038
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
```

```
C;Accession: A56038
R;Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.
Mol. Cell. Biol. 14, 6809-6818, 1994
A;Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaste
A;Reference number: A56038; MUID:95021209; PMID:7935398
A;Accession: A56038
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1028 <GAR>
A;Cross-references: GB:U11383; NID:g520526; PIDN:AAB60216.1; PID:g520527
C;Genetics:
A;Gene: ovo
A;Cross-references: FlyBase:FBgn0003028

  Query Match             63.3%;  Score 69;  DB 2;  Length 1028;
  Best Local Similarity   85.7%;  Pred. No. 4.2;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAA 23
                ||||| ||||| ||||| |||
Db        497 AAAAAAAAAAAAAAAAAAAAA 517

RESULT 8
S16356
ovo protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997
C;Accession: S16356
R;Mevel-Ninio, M.; Terracol, R.; Kafatos, F.C.
EMBO J. 10, 2259-2266, 1991
A;Title: The ovo gene of Drosophila encodes a zinc finger protein required for female ge
A;Reference number: S16356; MUID:91293102; PMID:1712294
A;Accession: S16356
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1213 <MEV>
A;Cross-references: EMBL:X59772
C;Genetics:
A;Gene: FlyBase:ovo
A;Cross-references: FlyBase:FBgn0003028
A;Introns: 931/3; 1152/3

  Query Match             63.3%;  Score 69;  DB 2;  Length 1213;
  Best Local Similarity   85.7%;  Pred. No. 4.7;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAA 23
                ||||| ||||| ||||| |||
Db        860 AAAAAAAAAAAAAAAAAAAAA 880

RESULT 9
T13748
sex comb protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13748
R;Sinclair, D.A.R.; Milne, T.A.; Hodgson, J.W.; Shellard, J.; Salinas, C.A.; Kyba, M.; R
Development 125, 1207-1216, 1998
A;Title: The Additional sex combs gene of Drosophila encodes a chromatin protein that bi
A;Reference number: Z17750; MUID:98146384; PMID:9477319
A;Accession: T13748
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1668 <SIN>
A;Cross-references: EMBL:AJ001164; NID:g3292938; PIDN:CAA04568.1; PID:g3292939
C;Genetics:
A;Cross-references: FlyBase:FBgn0000142
C;Function:
A;Description: involved in repression of homeotic loci

  Query Match             63.3%;  Score 69;  DB 2;  Length 1668;
  Best Local Similarity   78.3%;  Pred. No. 6;
  Matches   18;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAA 24
                :||||| ||||| ||||| || |
Db        128 KAAAAAAAAAAAAAAAAAAAAQA 150

RESULT 10
A81882
probable dihydrolipoamide S-succinyltransferase (EC 2.3.1.61) E2 component NMA1150 [impo
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: A81882
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: A81882
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-403 <PAR>
A;Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84412.1; PID:g737984
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: sucB; NMA1150
C;Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
C;Keywords: acyltransferase; coenzyme A

  Query Match             62.4%;  Score 68;  DB 2;  Length 403;
  Best Local Similarity   75.0%;  Pred. No. 2.7;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
                || ||| ||| | ||| |||| ||
Db         85 EAPAAATAAAEAPAAAPAEAAPAA 108

RESULT 11
A29945
neurogenesis regulatory protein - fruit fly (Drosophila melanogaster) (fragment)
N;Alternate names: single-minded gene protein
C;Species: Drosophila melanogaster
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 20-Mar-1998
C;Accession: A29945
R;Crews, S.T.; Thomas, J.B.; Goodman, C.S.
Cell 52, 143-151, 1988
A;Title: The Drosophila single-minded gene encodes a nuclear protein with sequence simil
A;Reference number: A29945; MUID:88151023; PMID:3345560
A;Accession: A29945
A;Molecule type: mRNA
A;Residues: 1-655 <CRE>
A;Cross-references: GB:M19020; NID:g158464; PID:g158465
C;Genetics:
A;Gene: sim
A;Cross-references: FlyBase:FBgn0004666
C;Keywords: DNA binding; transcription regulation

  Query Match             62.4%;  Score 68;  DB 2;  Length 655;
  Best Local Similarity   66.7%;  Pred. No. 3.8;
  Matches   16;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
                :|| ||:|| ||:|| ||:|| ||
Db        366 QAAQAAQAAQAAQAAQAAQAAQAA 389

RESULT 12
B53225
ecdysone-induced protein E74A - fruit fly (Drosophila virilis)
C;Species: Drosophila virilis
C;Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Feb-1997
```

```
C;Accession: B53225
R;Jones, C.W.; Dalton, M.W.; Townley, L.H.
Genetics 127, 535-543, 1991
A;Title: Interspecific comparisons of the structure and regulation of the Drosophila ecd
A;Reference number: A53225; MUID:91200627; PMID:2016053
A;Accession: B53225
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-873 <JON>
A;Cross-references: GB:X59493
C;Genetics:
A;Gene: FlyBase:Dvir/Eip74EF
A;Cross-references: FlyBase:FBgn0013076
C;Superfamily: ets DNA-binding domain homology
F;779-859/Domain: ets DNA-binding domain homology <ETS>

  Query Match             62.4%;  Score 68;  DB 2;  Length 873;
  Best Local Similarity   73.9%;  Pred. No. 4.7;
  Matches   17;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              |:||| ||||  :|||| |||||
Db        501 ASAAAAAAAATHSAAAAAAAAAA 523

RESULT 13
T13154
polycomb protein enhancer - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13154
R;Stankunas, K.; Berger, J.; Ruse, C.; Sinclair, D.A.; Randazzo, F.; Brock, H.W.
Development 125, 4055-4066, 1998
A;Title: The enhancer of polycomb gene of Drosophila encodes a chromatin protein conserv
A;Reference number: Z17611; MUID:98407961; PMID:9735366
A;Accession: T13154
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2023 <STA>
A;Cross-references: EMBL:AF079764; NID:g3757889; PID:g3757890; PIDN:AAC64271.1
A;Experimental source: imaginal disc
C;Genetics:
A;Gene: E(Pc)
A;Cross-references: FlyBase:FBgn0000581
A;Map position: 2

  Query Match             62.4%;  Score 68;  DB 2;  Length 2023;
  Best Local Similarity   60.0%;  Pred. No. 8.7;
  Matches   18;  Conservative    3;  Mismatches    3;  Indels    6;  Gaps    1;

Qy          1 CE------AAAAAEAAAAAEAAAAAEAAAA 24
              ||      :||| || ||||||||||::
Db        810 CEDQPVASTSAAAAAARAAEAAAAAEAASS 839

RESULT 14
JC5273
paired type homeobox protein, NBP - human
C;Species: Homo sapiens (man)
C;Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 24-Sep-1999
C;Accession: JC5273
R;Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.
DNA Res. 3, 311-320, 1996
A;Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific
A;Reference number: JC5272; MUID:97191543; PMID:9039501
A;Contents: neuroblastma cell
A;Accession: JC5273
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-314 <YOK>
A;Cross-references: DDBJ:D82344; NID:g1841337; PIDN:BAA11555.1; PID:d1012222; PID:g18413
C;Comment: This protein is a transcriptional repressor involved in regulating gene expre
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;99-155/Domain: homeobox homology <HOX>

  Query Match             61.5%;  Score 67;  DB 2;  Length 314;
  Best Local Similarity   78.3%;  Pred. No. 2.8;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              ||||| ||||| |||||   |||
Db        244 AAAAAAAAAAAAAAAAAGGLAAA 266

RESULT 15
T33110
hypothetical protein C18H7.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C;Accession: T33110
R;Tin-Wollam, A.; Fronick, W.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid C18H7.
A;Reference number: Z21284
A;Accession: T33110
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-460 <TIN>
A;Cross-references: EMBL:AF067607; PIDN:AAC17641.1; GSPDB:GN00022; CESP:C18H7.3
A;Experimental source: strain Bristol N2; clone C18H7
C;Genetics:
A;Gene: CESP:C18H7.3
A;Map position: 4
A;Introns: 84/1
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

  Query Match             61.5%;  Score 67;  DB 2;  Length 460;
  Best Local Similarity   62.5%;  Pred. No. 3.7;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    8;  Gaps    1;

Qy          2 EAAAAAEAAAAAE--------AAAAAEAAAAA 25
              ||||| ||| |||        | || ||||||
Db        399 EAAAAPEAAPAAEGAGGGAEPAGAAPEAAAAA 430

Search completed: January 30, 2004, 00:26:21
Job time : 11.6808 secs
```

```
OM protein - protein search, using sw model

Run on:         January 29, 2004, 23:58:27 ; Search time 5.6338 Seconds
                                             (without alignments)
                                             208.681 Million cell updates/sec

Title:          US-09-461-684C-2
Perfect score:  109
Sequence:       1 CEAAAAAEAAAAAEAAAAAEAAAAA 25

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79	72.5	332	1	FA9A_HUMAN	Q8izu1 homo sapien
2	75.5	69.3	518	1	TPM4_DROME	P49455 drosophila
3	72	66.1	165	1	RS16_CAUCR	P58122 caulobacter
4	70	64.2	183	1	OLEC_BRANA	P29526 brassica na
5	70	64.2	401	1	HME1_MOUSE	P09065 mus musculu
6	69	63.3	97	1	ANP_LIMFE	P09031 limanda fer
7	69	63.3	1028	1	OVO_DROME	P51521 drosophila
8	69	63.3	1669	1	ASX_DROME	Q9v727 drosophila
9	68	62.4	697	1	SIM_DROME	P05709 drosophila
10	68	62.4	1073	1	HR38_DROME	P49869 drosophila
11	67	61.5	314	1	PMXB_HUMAN	Q99453 homo sapien
12	67	61.5	314	1	PMXB_MOUSE	O35690 mus musculu
13	67	61.5	477	1	MAZ_HUMAN	P56270 homo sapien
14	67	61.5	606	1	HM1D_DROAN	P22544 drosophila
15	66	60.6	386	1	HXAD_MOUSE	Q62424 mus musculu
16	66	60.6	388	1	HXAD_HUMAN	P31271 homo sapien
17	65	59.6	85	1	ANP4_PSEAM	P02734 pseudopleur
18	65	59.6	91	1	ANPX_PSEAM	P07835 pseudopleur
19	65	59.6	276	1	SX21_HUMAN	Q9y651 homo sapien
20	65	59.6	280	1	SX21_CHICK	Q9w7r5 gallus gall
21	65	59.6	475	1	EVX2_MOUSE	P49749 mus musculu
22	65	59.6	476	1	EVX2_HUMAN	Q03828 homo sapien
23	65	59.6	644	1	BTD_DROME	Q24266 drosophila
24	64	58.7	109	1	RLA1_TRYCR	P26643 trypanosoma
25	64	58.7	392	1	HME1_HUMAN	Q05925 homo sapien
26	64	58.7	1533	1	PUM_DROME	P25822 drosophila
27	64	58.7	2038	1	FSH_DROME	P13709 drosophila
28	63	57.8	179	1	RL19_AGRT5	Q8ubz5 agrobacteri
29	63	57.8	376	1	FXE1_HUMAN	O00358 homo sapien
30	63	57.8	521	1	RUN2_HUMAN	Q13950 h runt-rela
31	63	57.8	562	1	ARX_HUMAN	Q96qs3 homo sapien
32	63	57.8	590	1	HMAA_DROME	P29555 drosophila
33	63	57.8	607	1	RUN2_MOUSE	Q08775 m runt-rela
34	63	57.8	1095	1	PIPA_DROME	P13217 drosophila
35	62	56.9	364	1	NK61_MESAU	Q60554 mesocricetu
36	62	56.9	365	1	NK61_MOUSE	Q99ma9 mus musculu
37	62	56.9	365	1	NK61_RAT	O35762 rattus norv
38	62	56.9	1355	1	SALM_DROME	P39770 drosophila
39	61	56.0	31	1	ANP3_PAGBO	P02732 pagothenia
40	61	56.0	91	1	ANPY_PSEAM	P23699 pseudopleur
41	61	56.0	153	1	RS16_BIFLO	Q8g7g1 bifidobacte
42	61	56.0	308	1	ABF1_DROME	P39413 drosophila
43	61	56.0	376	1	FXL2_HUMAN	P58012 homo sapien
44	61	56.0	421	1	PO41_MOUSE	P17208 mus musculu
45	61	56.0	423	1	PO41_HUMAN	Q01851 homo sapien

ALIGNMENTS

```
RESULT 1
FA9A_HUMAN
ID   FA9A_HUMAN     STANDARD;      PRT;   332 AA.
AC   Q8IZU1;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Protein FAM9A.
GN   FAM9A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RX   MEDLINE=22202142; PubMed=12213195;
RA   Martinez-Garay I., Jablonka S., Sutajova M., Steuernagel P., Gal A.,
RA   Kutsche K.;
RT   "A new gene family (FAM9) of low-copy repeats in Xp22.3 expressed
RT   exclusively in testis: implications for recombinations in this
RT   region.";
RL   Genomics 80:259-267(2002).
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- TISSUE SPECIFICITY: Expressed exclusively in testis.
CC   -!- SIMILARITY: Belongs to the FAM9 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF494343; AAN07162.1; -.
DR   Genew; HGNC:18403; FAM9A.
KW   Nuclear protein.
FT   DOMAIN      180    275       GLU-RICH.
FT   DOMAIN      194    214       POLY-ALA.
FT   DOMAIN      252    258       POLY-GLY.
SQ   SEQUENCE   332 AA;  37339 MW;  92F22EC36038229C CRC64;

  Query Match            72.5%;  Score 79;  DB 1;  Length 332;
  Best Local Similarity  83.3%;  Pred. No. 0.066;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              ||  |  |||||||||||||||||
Db        190 EAEEAEAAAAAAEAAAAAEAAAAA 213


RESULT 2
TPM4_DROME
ID   TPM4_DROME     STANDARD;      PRT;   518 AA.
AC   P49455; P49456; Q24425; Q24426;
DT   01-FEB-1996 (Rel. 33, Created)
```

```
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Tropomyosin 1, isoforms 33/34 (Tropomyosin II).
GN   TM1 OR TMII.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC   TISSUE=Embryo, and Pupae;
RX   MEDLINE=89127197; PubMed=2851721;
RA   Hanke P.D., Storti R.V.;
RT   "The Drosophila melanogaster tropomyosin II gene produces multiple
RT   proteins by use of alternative tissue-specific promoters and
RT   alternative splicing.";
RL   Mol. Cell. Biol. 8:3591-3602(1988).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC   STRAIN=Oregon-R; TISSUE=Pupae;
RX   MEDLINE=87064486; PubMed=3097506;
RA   Karlik C.C., Fyrberg E.A.;
RT   "Two Drosophila melanogaster tropomyosin genes: structural and
RT   functional aspects.";
RL   Mol. Cell. Biol. 6:1965-1973(1986).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Additional isoforms seem to exist;
CC       Name=33; Synonyms=9C;
CC         IsoId=P49455-1; Sequence=Displayed;
CC       Name=Muscle; Synonyms=9D;
CC         IsoId=P06754-1; Sequence=External;
CC       Name=Non-muscle; Synonyms=Cytoskeletal;
CC         IsoId=P06754-2; Sequence=External;
CC       Name=9A;
CC         IsoId=P06754-3; Sequence=External;
CC       Name=34; Synonyms=9B;
CC         IsoId=P49455-2; Sequence=VSP_006623, VSP_006624, VSP_006625;
CC   -!- TISSUE SPECIFICITY: Both isoforms are only expressed in indirect
CC       flight muscles.
CC   -!- DEVELOPMENTAL STAGE: Both isoforms are expressed during pupal and
CC       adult stages.
CC   -!- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC       EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC   -!- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X76208; CAA53800.1; -.
DR   EMBL; X76208; CAA53801.1; -.
DR   EMBL; K02620; AAA28967.1; ALT_SEQ.
DR   EMBL; L00355; AAA28967.1; JOINED.
DR   EMBL; L00356; AAA28967.1; JOINED.
DR   EMBL; L00357; AAA28967.1; JOINED.
DR   EMBL; L00358; AAA28967.1; JOINED.
DR   EMBL; L00359; AAA28967.1; JOINED.
DR   EMBL; L00360; AAA28967.1; JOINED.
DR   EMBL; L00362; AAA28967.1; JOINED.
DR   EMBL; M12840; AAA28967.1; JOINED.
DR   EMBL; K02621; AAA28968.1; -.
DR   EMBL; M12840; AAA28968.1; JOINED.
DR   EMBL; L00355; AAA28968.1; JOINED.
DR   EMBL; L00356; AAA28968.1; JOINED.
DR   EMBL; L00357; AAA28968.1; JOINED.
DR   EMBL; L00358; AAA28968.1; JOINED.
DR   EMBL; L00359; AAA28968.1; JOINED.
DR   EMBL; L00360; AAA28968.1; JOINED.
DR   EMBL; L00362; AAA28968.1; JOINED.
DR   FlyBase; FBgn0003721; Tm1.
DR   InterPro; IPR000533; Tropomyosin.
DR   Pfam; PF00261; Tropomyosin; 1.
DR   PRINTS; PR00194; TROPOMYOSIN.
DR   PROSITE; PS00326; TROPOMYOSIN; 1.
KW   Muscle protein; Cytoskeleton; Actin-binding; Coiled coil;
KW   Alternative splicing; Multigene family.
FT   DOMAIN       14    267       COILED COIL (POTENTIAL).
FT   DOMAIN      287    518       ALA/PRO-RICH.
FT   VARSPLIC    259    293       DLIVEKERYCMIGDSLDEAFVDLIKGLEPFWNPRN -> EM
FT                                IKEIEHYALVGDQLDWTFVEMMGMPPFYNERY (in
FT                                isoform 34).
FT                                /FTId=VSP_006623.
FT   VARSPLIC    300    367       KLPTPTPEELAAMEEARAAAEAAAAAEAEAAEAAAAAGEAG
FT                                ADGAPAAPGEEKAPAKEPTPPKEPTPP -> ELTEEEKAAL
FT                                EAAAIAEKARAEELAALGEEAGAEAGEGGAPAEGAAPGEPG
FT                                AATEPGVEAPPAEPERIPT (in isoform 34).
FT                                /FTId=VSP_006624.
FT   VARSPLIC    391    518       KNYEPPPPGSEPEPVPAAEGEAAPAAEGAAPPAEGAAPPAE
FT                                GAVPPADGAAPPAEGAAPAAEGAAPPADGAAPPAEAAAAPA
FT                                DAAAPAAEAAPAEAPAAEATAAEAPPAEAAPAEAAPAAAEG
FT                                EAPPA -> RNAEPGDFAPPAEAAPAEGAPPAEGAPAAEGA
FT                                APAEGAPAAEGAPPAEGAPAPAPAEGEAAPPAPAAEGDAAA
FT                                APPPPAEGEAAPAPAEGEAPPAPPAEAAPAAEAPPA (in
FT                                isoform 34).
FT                                /FTId=VSP_006625.
FT   CONFLICT    106    114       LGSATAKLS -> SASAIQLAA (IN REF. 2).
FT   CONFLICT    119    119       A -> S (IN REF. 2).
FT   CONFLICT    183    183       A -> AMVEADLERAEERA (IN REF. 2).
FT   CONFLICT    199    199       V -> L (IN REF. 2).
FT   CONFLICT    503    503       P -> A (IN REF. 2; AAA28968).
SQ   SEQUENCE   518 AA;  54558 MW;  153D0872CF9DB6EA CRC64;

  Query Match             69.3%;  Score 75.5;  DB 1;  Length 518;
  Best Local Similarity   87.5%;  Pred. No. 0.21;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              || |||||||||| | ||||||||
Db        314 EARAAAEAAAAAE-AEAAEAAAAA 336


RESULT 3
RS16_CAUCR
ID   RS16_CAUCR     STANDARD;      PRT;   165 AA.
AC   P58122;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   30S ribosomal protein S16.
GN   RPSP OR CC3652.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC   -!- SIMILARITY: BELONGS TO THE S16P FAMILY OF RIBOSOMAL PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE006023; AAK25614.1; -.
DR   PIR; B87702; B87702.
DR   HSSP; P80379; 1EMW.
DR   TIGR; CC3652; -.
DR   HAMAP; MF_00385; -; 1.
DR   InterPro; IPR000307; Ribosomal_S16.
DR   Pfam; PF00886; Ribosomal_S16; 1.
DR   ProDom; PD003791; Ribosomal_S16; 1.
DR   TIGRFAMs; TIGR00002; S16; 1.
DR   PROSITE; PS00732; RIBOSOMAL_S16; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   165 AA;  17605 MW;  ED46FC2798C5BE1C CRC64;

  Query Match             66.1%;  Score 72;  DB 1;  Length 165;
  Best Local Similarity   75.0%;  Pred. No. 0.19;
  Matches   18;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              :| | |:||| || |||||||||
Db        115 QAEADAKAAAEAEKAAAAEAAAAA 138

RESULT 4
OLEC_BRANA
ID   OLEC_BRANA     STANDARD;      PRT;   183 AA.
AC   P29526;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Oleosin C98 (Fragment).
GN   C98.
OS   Brassica napus (Rape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Brassica.
OX   NCBI_TaxID=3708;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Anther;
RX   MEDLINE=93386188; PubMed=8374615;
RA   Roberts M.R., Hodge R., Ross J.H.E., Sorensen A., Murphy D.J.,
RA   Draper J., Scott R.;
RT   "Characterization of a new class of oleosins suggests a male
RT   gametophyte-specific lipid storage pathway.";
RL   Plant J. 3:629-636(1993).
CC   -!- FUNCTION: MAY HAVE A STRUCTURAL ROLE TO STABILIZE THE LIPID BODY
CC       DURING DESICCATION OF THE SEED BY PREVENTING COALESCENCE OF THE
CC       OIL. PROBABLY INTERACTS WITH BOTH LIPID AND PHOSPHOLIPID MOIETIES
CC       OF LIPID BODIES. MAY ALSO PROVIDE RECOGNITION SIGNALS FOR SPECIFIC
CC       LIPASE ANCHORAGE IN LIPOLYSIS DURING SEEDLING GROWTH.
CC   -!- SUBCELLULAR LOCATION: SURFACE OF OIL BODIES. OLEOSINS EXIST AT A
CC       MONOLAYER LIPID/WATER INTERFACE.
CC   -!- TISSUE SPECIFICITY: SPECIFIC TO THE MALE GAMETOPHYTE.
CC   -!- SIMILARITY: BELONGS TO THE OLEOSIN FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X67142; CAA47623.1; -.
DR   PIR; S24960; S24960.
DR   InterPro; IPR000136; Oleosin.
DR   Pfam; PF01277; Oleosin; 1.
DR   PROSITE; PS00811; OLEOSINS; 1.
KW   Seed; Oil body; Multigene family.
FT   NON_TER       1      1
FT   DOMAIN       <1     23       POLAR.
FT   DOMAIN       24     95       HYDROPHOBIC.
SQ   SEQUENCE   183 AA;  18149 MW;  198A5D3B6DF3045A CRC64;

  Query Match             64.2%;  Score 70;  DB 1;  Length 183;
  Best Local Similarity   78.3%;  Pred. No. 0.33;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              || ||| | ||||| ||||| ||
Db        151 AAPAAEPAPAAEAAPAAEAAPAA 173

RESULT 5
HME1_MOUSE
ID   HME1_MOUSE     STANDARD;      PRT;   401 AA.
AC   P09065;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein engrailed-1 (Mo-En-1).
GN   EN1 OR EN-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93185339; PubMed=1363401;
RA   Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
RA   Provart N.J., Joyner A.L.;
RT   "Cloning and sequence comparison of the mouse, human, and chicken
RT   engrailed genes reveal potential functional domains and regulatory
RT   regions.";
RL   Dev. Genet. 13:345-358(1992).
RN   [2]
RP   SEQUENCE OF 278-401 FROM N.A.
RX   MEDLINE=88112776; PubMed=2892757;
RA   Joyner A.L., Martin G.R.;
RT   "En-1 and En-2, two mouse genes with sequence homology to the
RT   Drosophila engrailed gene: expression during embryogenesis.";
RL   Genes Dev. 1:29-38(1987).
RN   [3]
RP   SEQUENCE OF 298-401 FROM N.A.
RX   MEDLINE=86079501; PubMed=2416459;
RA   Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;
RT   "Expression during embryogenesis of a mouse gene with sequence
RT   homology to the Drosophila engrailed gene.";
RL   Cell 43:29-37(1985).
RN   [4]
RP   SEQUENCE OF 321-380 FROM N.A.
RX   MEDLINE=91099509; PubMed=1980115;
RA   Holland P.W.H., Williams N.A.;
RT   "Conservation of engrailed-like homeobox sequences during vertebrate
RT   evolution.";
RL   FEBS Lett. 277:250-252(1990).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE ENGRAILED HOMEOBOX FAMILY.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L12703; AAA03660.2; -.
```

```
DR   EMBL; Y00201; CAA68361.1; -.
DR   PIR; A48423; A48423.
DR   HSSP; P02836; 3HDD.
DR   TRANSFAC; T02016; -.
DR   MGD; MGI:95389; En1.
DR   InterPro; IPR000747; Engrailed.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR000047; HTH_lambrepressr.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00026; ENGRAILED.
DR   PRINTS; PR00024; HOMEOBOX.
DR   PRINTS; PR00031; HTHREPRESSR.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
DR   PROSITE; PS00033; ENGRAILED; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DOMAIN       52     87       PRO-RICH.
FT   DOMAIN       73     87       POLY-PRO.
FT   DOMAIN      207    228       POLY-ALA.
FT   DNA_BIND    312    371       HOMEOBOX.
SQ   SEQUENCE   401 AA;  40950 MW;  1F90210950152FAE CRC64;

  Query Match             64.2%;  Score 70;  DB 1;  Length 401;
  Best Local Similarity   78.3%;  Pred. No. 0.62;
  Matches   18;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAAEAAAAA 25
              ||||| ||||| || || ||||:
Db        207 AAAAAAAAAAAAAAVAAAAAAAS 229


RESULT 6
ANP_LIMFE
ID   ANP_LIMFE      STANDARD;      PRT;    97 AA.
AC   P09031;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-AUG-1990 (Rel. 15, Last annotation update)
DE   Antifreeze protein precursor (AFP).
OS   Limanda ferruginea (Yellowtail flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Pleuronectidae; Limanda.
OX   NCBI_TaxID=8258;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88029483; PubMed=3665937;
RA   Scott G.K., Davies P.L., Shears M.A., Fletcher G.L.;
RT   "Structural variations in the alanine-rich antifreeze proteins of the
RT   pleuronectinae.";
RL   Eur. J. Biochem. 168:629-633(1987).
CC   -!- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC   -!- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE
CC       ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X06356; CAA29655.1; -.
DR   PIR; S02376; S02376.
DR   InterPro; IPR000104; Antifreeze_1.
DR   PRINTS; PR00308; ANTIFREEZEI.
KW   Antifreeze protein; Repeat; Signal.
FT   SIGNAL        1     23
FT   PROPEP       24     48       REMOVED BY A DIPEPTIDYLPEPTIDASE
FT                                (PROBABLE).
FT   CHAIN        49     97       ANTIFREEZE PROTEIN.
SQ   SEQUENCE   97 AA;  8865 MW;  62AD582DF8E459B6 CRC64;

  Query Match             63.3%;  Score 69;  DB 1;  Length 97;
  Best Local Similarity   70.8%;  Pred. No. 0.26;
  Matches   17;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAAEAAAAA 25
              :||||| | ||| | |||: ||||
Db         53 DAAAAAAATAAAAAKAAADTAAAA 76


RESULT 7
OVO_DROME
ID   OVO_DROME      STANDARD;      PRT;  1028 AA.
AC   P51521; Q9XZU4;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Ovo protein (Shaven baby protein).
GN   OVO OR SVB.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=95021209; PubMed=7935398;
RA   Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;
RT   "Multiple products from the shavenbaby-ovo gene region of Drosophila
RT   melanogaster: relationship to genetic complexity.";
RL   Mol. Cell. Biol. 14:6809-6818(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R;
RX   MEDLINE=91293102; PubMed=1712294;
RA   Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;
RT   "The ovo gene of Drosophila encodes a zinc finger protein required
RT   for female germ line development.";
RL   EMBO J. 10:2259-2266(1991).
CC   -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM
CC       LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIUM AND
CC       ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,
CC       BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED
CC       PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.
CC   -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U11383; AAB60216.1; -.
DR   EMBL; X59772; CAB36921.1; ALT_SEQ.
DR   PIR; A56038; A56038.
DR   HSSP; P07248; 2ADR.
DR   TRANSFAC; T00669; -.
DR   FlyBase; FBgn0003028; ovo.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 3.
DR   SMART; SM00355; ZnF_C2H2; 4.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW   Transcription regulation.
```

```
FT   DOMAIN       62     66       POLY-ALA.
FT   DOMAIN       72     77       POLY-GLY.
FT   DOMAIN       80     85       POLY-GLY.
FT   DOMAIN       98    108       POLY-GLY.
FT   DOMAIN      144    152       POLY-HIS.
FT   DOMAIN      153    159       POLY-ASN.
FT   DOMAIN      336    339       POLY-GLN.
FT   DOMAIN      347    353       POLY-GLN.
FT   DOMAIN      357    361       POLY-GLN.
FT   DOMAIN      410    414       POLY-GLN.
FT   DOMAIN      418    422       POLY-GLN.
FT   DOMAIN      426    432       POLY-GLN.
FT   DOMAIN      445    453       POLY-GLN.
FT   DOMAIN      456    459       POLY-GLN.
FT   DOMAIN      466    474       POLY-GLN.
FT   DOMAIN      497    517       POLY-ALA.
FT   DOMAIN      524    529       POLY-SER.
FT   DOMAIN      549    558       POLY-ALA.
FT   DOMAIN      639    651       POLY-ALA.
FT   DOMAIN      717    725       POLY-ALA.
FT   DOMAIN      797    802       POLY-GLN.
FT   DOMAIN      820    823       POLY-GLN.
FT   DOMAIN      826    832       POLY-GLN.
FT   ZN_FING     874    896       C2H2-TYPE 1.
FT   ZN_FING     902    924       C2H2-TYPE 2.
FT   ZN_FING     930    953       C2H2-TYPE 3.
FT   ZN_FING     969    992       C2H2-TYPE 4.
FT   CONFLICT    647    647       A -> R (IN REF. 2).
SQ   SEQUENCE   1028 AA;  110620 MW;  D7068BB2BC0F6F77 CRC64;

  Query Match             63.3%;  Score 69;  DB 1;  Length 1028;
  Best Local Similarity   85.7%;  Pred. No. 1.6;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAA 23
              ||||| ||||| ||||| |||
Db        497 AAAAAAAAAAAAAAAAAAAAA 517


RESULT 8
ASX_DROME
ID   ASX_DROME      STANDARD;      PRT;  1669 AA.
AC   Q9V727; O76930;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Polycomb protein Asx (Additional sex combs).
GN   ASX OR CG8787.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, AND DEVELOPMENTAL
RP   STAGE.
RC   TISSUE=Imaginal disks;
RX   MEDLINE=98146384; PubMed=9477319;
RA   Sinclair D.A.R., Milne T.A., Hodgson J.W., Shellard J., Salinas C.A.,
RA   Kyba M., Randazzo F., Brock H.W.;
RT   "The Additional sex combs gene of Drosophila encodes a chromatin
RT   protein that binds to shared and unique Polycomb group sites on
RT   polytene chromosomes.";
RL   Development 125:1207-1216(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   INTERACTION WITH TAN.
RX   MEDLINE=21290825; PubMed=11397012;
RA   Dietrich B.H., Moore J., Kyba M., dosSantos G., McCloskey F.,
RA   Milne T.A., Brock H.W., Krause H.M.;
RT   "Tantalus, a novel ASX-interacting protein with tissue-specific
RT   functions.";
RL   Dev. Biol. 234:441-453(2001).
CC   -!- FUNCTION: Atypical Polycomb group protein, which may be involved
CC       in both Polycomb group (PcG) and trithorax group (trxG) complexes.
CC       PcG and trxG proteins act by forming multiprotein complexes, which
CC       are respectively required to maintain the transcriptionally
CC       repressive and transcriptionally active state of homeotic genes
CC       throughout development. PcG and trxG protein complexes are not
CC       required to initiate repression and activation, but to maintain it
CC       during later stages of development. Both complexes probably act
CC       via methylation of histones, rendering chromatin heritably changed
CC       in its expressibility.
CC   -!- SUBUNIT: Interacts with Tan.
CC   -!- SUBCELLULAR LOCATION: Nuclear; associated with chromatin.
CC       Colocalizes with many PcG sites on polytene chromosomes. It also
CC       associates with many unique sites on polytene chromosomes.
CC   -!- TISSUE SPECIFICITY: Highly expressed in nurse cells and deposited
CC       in oocytes late in oogenesis. Ubiquitous in early embryos. Late
CC       embryos show higher levels in CNS and neurectoderm.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC       Early embryos have high levels of expression, this drops off and
CC       zygotic expression begins at 3-6 hour embryos. Expression levels
CC       are low in larvae and medium in pupae and adults.
CC   -!- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs, which may
CC       be required for an association with nuclear receptors (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the Asx family.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to
CC       frameshifts in positions 608 and 719.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AJ001164; CAA04568.1; ALT_FRAME.
DR   EMBL; AE003814; AAF58239.1; -.
DR   FlyBase; FBgn0000142; Asx.
DR   PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR   PROSITE; PS50016; ZF_PHD_2; FALSE_NEG.
KW   Transcription regulation; Repressor; Nuclear protein; Zinc;
KW   Metal-binding; Zinc-finger; Repeat; Developmental protein.
FT   ZN_FING     1632   1669       PHD-TYPE (ATYPICAL).
FT   DOMAIN         8     12       POLY-THR.
FT   DOMAIN       122    126       POLY-GLN.
FT   DOMAIN       129    152       ALA-RICH.
FT   DOMAIN       638    715       SER-RICH.
FT   DOMAIN       747    751       POLY-GLN.
FT   DOMAIN       862   1202       GLN-RICH.
FT   DOMAIN      1287   1290       POLY-THR.
FT   DOMAIN      1520   1524       POLY-HIS.
FT   DOMAIN      1527   1536       POLY-GLN.
FT   SITE         224    228       LXXLL MOTIF 1.
FT   SITE         244    248       LXXLL MOTIF 2.
FT   CONFLICT      14     15       SQ -> CE (IN REF. 1).
FT   CONFLICT     187    187       K -> N (IN REF. 1).
FT   CONFLICT    1253   1253       S -> T (IN REF. 1).
FT   CONFLICT    1520   1520       MISSING (IN REF. 1).
SQ   SEQUENCE   1669 AA;  179841 MW;  F65D87398D67D321 CRC64;

  Query Match             63.3%;  Score 69;  DB 1;  Length 1669;
  Best Local Similarity   78.3%;  Pred. No. 2.4;
  Matches   18;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAA 24
              :||||| ||||| ||||| || |
Db        128 KAAAAAAAAAAAAAAAAAAAAQA 150

RESULT 9
SIM_DROME
ID   SIM_DROME      STANDARD;      PRT;   697 AA.
AC   P05709; O96521; Q8MQI7; Q9VFZ3;
DT   01-NOV-1988 (Rel. 09, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Single-minded protein.
GN   SIM OR CG7771.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99054545; PubMed=9840810;
RA   Kasai Y., Stahl S., Crews S.;
RT   "Specification of the Drosophila CNS midline cell lineage: direct
RT   control of single-minded transcription by dorsal/ventral patterning
RT   genes.";
RL   Gene Expr. 7:171-189(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS.
RC   STRAIN=Berkeley;
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN   [5]
RP   SEQUENCE OF 25-42 FROM N.A., AND SIMILARITY TO HLH PROTEINS.
RX   MEDLINE=92103681; PubMed=1760843;
RA   Nambu J.R., Lewis J.O., Wharton K.A. Jr., Crews S.T.;
RT   "The Drosophila single-minded gene encodes a helix-loop-helix protein
RT   that acts as a master regulator of CNS midline development.";
RL   Cell 67:1157-1167(1991).
RN   [6]
RP   SEQUENCE OF 43-697 FROM N.A.
RX   MEDLINE=88151023; PubMed=3345560;
RA   Crews S.T., Thomas J.B., Goodman C.S.;
RT   "The Drosophila single-minded gene encodes a nuclear protein with
RT   sequence similarity to the per gene product.";
RL   Cell 52:143-151(1988).
CC   -!- FUNCTION: TRANSCRIPTION FACTOR THAT FUNCTIONS AS A MASTER
CC       DEVELOPMENTAL REGULATOR OF THE CNS MIDLINE LINEAGE. MUTATIONS IN
CC       THE SIM GENE RESULTS IN THE LOSS OF THE PRECURSOR CELLS GIVING
CC       RISE TO MIDLINE CELLS OF THE EMBRYONIC CENTRAL NERVOUS SYSTEM.
CC   -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC       BHLH protein.
```

```
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- POLYMORPHISM: Berkeley strain has 11 A-A-Q repeats.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS.
CC   -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
CC   -!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC       gene model prediction.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF071934; AAC64519.1; ALT_SEQ.
DR   EMBL; AE003698; AAF54902.3; -.
DR   EMBL; AY129457; AAM76199.1; -.
DR   EMBL; M19020; AAA28900.1; -.
DR   PIR; A29945; A29945.
DR   TRANSFAC; T00750; -.
DR   FlyBase; FBgn0004666; sim.
DR   GO; GO:0005634; C:nucleus; IEP.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0007418; P:ventral midline development; IMP.
DR   InterPro; IPR001092; HLH_basic.
DR   InterPro; IPR001067; Nuc_translocat.
DR   InterPro; IPR001610; PAC.
DR   InterPro; IPR000014; PAS_domain.
DR   Pfam; PF00010; HLH; 1.
DR   Pfam; PF00785; PAC; 1.
DR   Pfam; PF00989; PAS; 1.
DR   PRINTS; PR00785; NCTRNSLOCATR.
DR   SMART; SM00353; HLH; 1.
DR   SMART; SM00086; PAC; 1.
DR   SMART; SM00091; PAS; 2.
DR   TIGRFAMs; TIGR00229; sensory_box; 2.
DR   PROSITE; PS00038; HLH_1; 1.
DR   PROSITE; PS50888; HLH_2; 1.
DR   PROSITE; PS50112; PAS; 2.
KW   Developmental protein; Neurogenesis; Nuclear protein; Repeat;
KW   Transcription regulation; DNA-binding.
FT   DNA_BIND     21     37       BASIC DOMAIN.
FT   DOMAIN       38     78       HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT   DOMAIN      100    172       PAS 1.
FT   DOMAIN      266    336       PAS 2.
FT   DOMAIN      406    446       14 X 3 AA REPEATS OF A-A-Q (APPROXIMATE).
FT   DOMAIN      553    672       SER-RICH.
FT   DOMAIN      673    693       GLN/HIS-RICH.
FT   VARIANT     406    414       MISSING (IN STRAIN BERKELEY).
FT   CONFLICT    151    151       I -> Y (IN REF. 1).
SQ   SEQUENCE   697 AA;  76475 MW;  588414A4A17101AD CRC64;

  Query Match            62.4%;  Score 68;  DB 1;  Length 697;
  Best Local Similarity   66.7%;  Pred. No. 1.5;
  Matches   16;  Conservative    4;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              :|| ||:|| ||:|| ||:|| ||
Db        408 QAAQAAQAAQAAQAAQAAQAAQAA 431


RESULT 10
HR38_DROME
ID   HR38_DROME     STANDARD;      PRT;  1073 AA.
AC   P49869; O18383; Q9VIK4;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable nuclear hormone receptor HR38 (dHR38).
GN   HR38 OR NR4A4 OR CG1864.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   TISSUE=Larva;
RX   MEDLINE=95372400; PubMed=7644522;
RA   Sutherland J.D., Kozlova T., Tzertzinis G., Kafatos F.C.;
RT   "Drosophila hormone receptor 38: a second partner for Drosophila USP
RT   suggests an unexpected role for nuclear receptors of the nerve growth
RT   factor-induced protein B type.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:7966-7970(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT), AND TISSUE SPECIFICITY.
RX   MEDLINE=98370123; PubMed=9704500;
RA   Komonyi O., Mink M., Csiha J., Maroy P.;
RT   "Genomic organization of DHR38 gene in Drosophila: presence of
RT   Alu-like repeat in a translated exon and expression during embryonic
RT   development.";
RL   Arch. Insect Biochem. Physiol. 38:185-192(1998).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM LONG), AND CHARACTERIZATION.
RC   TISSUE=Larva;
RX   MEDLINE=98315108; PubMed=9649534;
RA   Kozlova T., Pokholkova G.V., Tzertzinis G., Sutherland J.D.,
RA   Zhimulev I.F., Kafatos F.C.;
RT   "Drosophila hormone receptor 38 functions in metamorphosis: a role in
RT   adult cuticle formation.";
RL   Genetics 149:1465-1475(1998).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM LONG).
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
```

```
RL   Science 287:2185-2195(2000).
RN   [5]
RP   SEQUENCE OF 528-1073 FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE=96068664; PubMed=7479849;
RA   Fisk G.J., Thummel C.S.;
RT   "Isolation, regulation, and DNA-binding properties of three
RT   Drosophila nuclear hormone receptor superfamily members.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:10604-10608(1995).
CC   -!- FUNCTION: BINDS TO NGFI-B RESPONSE ELEMENTS. PLAYS AN IMPORTANT
CC       ROLE IN LATE STAGES OF EPIDERMAL METAMORPHOSIS.
CC   -!- SUBUNIT: FORMS A HETERODIMER WITH USP.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=P49869-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=P49869-2; Sequence=VSP_003714;
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN PREBLASTODERM
CC       EMBRYOS, SPECIFICALLY IN CENTRAL NERVOUS SYSTEM AND INTESTINAL
CC       TRACT. HIGHLY EXPRESSED IN THIRD INSTAR LARVAL IMAGINAL DISKS AND
CC       BRAIN COMPLEXES, BUT NOT IN OVARIES.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN 0-8 HOUR EMBRYOS AND ADULTS.
CC       HIGHER IN LATE EMBRYOGENESIS AND DURING LARVAL AND PUPAL STAGES.
CC       SHORT ISOFORM IS ENRICHED IN PUPAE AND ADULTS, LONG ISOFORM IN
CC       LARVAE.
CC   -!- SIMILARITY: Belongs to the nuclear hormone receptor family. NR4
CC       subfamily.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X89246; CAA61534.1; -.
DR   EMBL; Y15606; CAA75690.1; -.
DR   EMBL; AJ002073; CAA05172.1; -.
DR   EMBL; AE003667; AAF53914.1; -.
DR   EMBL; U36762; AAC46926.1; -.
DR   HSSP; P19793; 2NLL.
DR   TRANSFAC; T02760; -.
DR   FlyBase; FBgn0014859; Hr38.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Stdhrmn_receptor.
DR   InterPro; IPR001628; Znf_C4steroid.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   ProDom; PD000035; Znf_C4steroid; 1.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger; Alternative splicing; Developmental protein.
FT   DNA_BIND     744    809       NUCLEAR RECEPTOR-TYPE.
FT   ZN_FING      744    764       C4-TYPE.
FT   ZN_FING      780    804       C4-TYPE.
FT   DOMAIN       188    192       POLY-ALA.
FT   DOMAIN       206    218       POLY-ALA.
FT   DOMAIN       221    228       POLY-ALA.
FT   DOMAIN       268    272       POLY-THR.
FT   DOMAIN       294    312       POLY-GLN.
FT   DOMAIN       441    462       POLY-GLN.
FT   DOMAIN       505    508       POLY-SER.
FT   DOMAIN       619    626       POLY-GLN.
FT   DOMAIN       661    665       POLY-ALA.
FT   VARSPLIC       1    522       Missing (in isoform Short).
FT                                 /FTId=VSP_003714.
FT   CONFLICT     667    667       V -> VSSPSV (IN REF. 4).
FT   CONFLICT     685    685       S -> L (IN REF. 1 AND 3).
FT   CONFLICT     689    692       STAQ -> LHGER (IN REF. 2).
FT   CONFLICT     697    697       A -> D (IN REF. 2).
FT   CONFLICT     702    702       N -> S (IN REF. 2).
FT   CONFLICT    1041   1041       S -> R (IN REF. 2).
FT   CONFLICT    1064   1064       E -> D (IN REF. 2).
SQ   SEQUENCE   1073 AA;  116991 MW;  126A30DAFA1C096A CRC64;

  Query Match             62.4%;  Score 68;  DB 1;  Length 1073;
  Best Local Similarity   73.9%;  Pred. No. 2.1;
  Matches   17;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              || ||  ||||||  || |||||
Db        206 AATAAATAAAAEAGGAASAAAAA 228


RESULT 11
PMXB_HUMAN
ID   PMXB_HUMAN     STANDARD;      PRT;   314 AA.
AC   Q99453;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
DE   (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).
GN   PHOX2B OR PMX2B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neuroblastoma;
RX   MEDLINE=97191543; PubMed=9039501;
RA   Yokoyama M., Nishi Y., Yoshii J., Okubo K., Matsubara K.;
RT   "Identification and cloning of neuroblastoma-specific and nerve
RT   tissue-specific genes through compiled expression profiles.";
RL   DNA Res. 3:311-320(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20487360; PubMed=11034547;
RA   Adachi M., Browne D., Lewis E.J.;
RT   "Paired-like homeodomain proteins Phox2a/Arix and Phox2b/NBPhox have
RT   similar genetic organization and independently regulate dopamine
RT   beta-hydroxylase gene transcription.";
RL   DNA Cell Biol. 19:539-554(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99326521; PubMed=10395798;
RA   Yokoyama M., Watanabe H., Nakamura M.;
RT   "Genomic structure and functional characterization of NBPhox (PMX2B),
RT   a homeodomain protein specific to catecholaminergic cells that is
RT   involved in second messenger-mediated transcriptional activation.";
RL   Genomics 59:40-50(1999).
CC   -!- FUNCTION: INVOLVED IN THE DEVELOPMENT OF SEVERAL MAJOR
CC       NORADRENERGIC NEURON POPULATIONS, INCLUDING THE LOCUS COERULEUS.
CC       TRANSCRIPTION FACTOR WHICH COULD DETERMINE A NEUROTRANSMITTER
CC       PHENOTYPE IN VERTEBRATES. ENHANCES SECOND-MESSENGER-MEDIATED
CC       ACTIVATION OF THE DOPAMINE BETA-HYDROYLASE AND C-FOS PROMOTERS,
CC       AND OF SEVERAL ENHANCERS INCLUDING CYCLIC AMP-RESPONSE ELEMENT AND
CC       SERUM-RESPONSE ELEMENT.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN NEUROBLASTOMA, BRAIN AND ADRENAL
CC       GLAND.
CC   -!- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D82344; BAA11555.1; -.
DR   EMBL; AF117979; AAD26698.1; -.
DR   EMBL; AB015671; BAA82670.1; -.
DR   PIR; JC5273; JC5273.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T03961; -.
DR   Genew; HGNC:9143; PHOX2B.
DR   MIM; 603851; -.
DR   GO; GO:0003712; F:transcription cofactor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW   Transcription regulation; Activator.
FT   DNA_BIND     98    157       HOMEOBOX.
FT   DOMAIN      159    167       POLY-ALA.
FT   DOMAIN      212    217       POLY-GLY.
FT   DOMAIN      241    260       POLY-ALA.
SQ   SEQUENCE   314 AA;  31607 MW;  76737F71948B5D81 CRC64;

  Query Match             61.5%;  Score 67;  DB 1;  Length 314;
  Best Local Similarity   78.3%;  Pred. No. 1;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              ||||| ||||| |||||   |||
Db        244 AAAAAAAAAAAAAAAAAGGLAAA 266


RESULT 12
PMXB_MOUSE
ID   PMXB_MOUSE     STANDARD;      PRT;   314 AA.
AC   O35690;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Paired mesoderm homeobox protein 2B (Paired-like homeobox 2B)
DE   (PHOX2B homeodomain protein) (Neuroblastoma Phox) (NBPhox).
GN   PHOX2B OR PMX2B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98040559; PubMed=9374403;
RA   Pattyn A., Morin X., Cremer H., Goridis C., Brunet J.-F.;
RT   "Expression and interactions of the two closely related homeobox
RT   genes Phox2a and Phox2b during neurogenesis.";
RL   Development 124:4065-4075(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99326521; PubMed=10395798;
RA   Yokoyama M., Watanabe H., Nakamura M.;
RT   "Genomic structure and functional characterization of NBPhox (PMX2B),
RT   a homeodomain protein specific to catecholaminergic cells that is
RT   involved in second messenger-mediated transcriptional activation.";
RL   Genomics 59:40-50(1999).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y14493; CAA74833.1; -.
DR   EMBL; AB015672; BAA82671.1; -.
DR   HSSP; P06601; 1FJL.
DR   TRANSFAC; T03976; -.
DR   MGD; MGI:1100882; Phox2b.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR007104; Paired_homeo.
DR   Pfam; PF00046; homeobox; 1.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT   DNA_BIND     98    157       HOMEOBOX.
FT   DOMAIN      159    167       POLY-ALA.
FT   DOMAIN      212    217       POLY-GLY.
FT   DOMAIN      241    260       POLY-ALA.
SQ   SEQUENCE   314 AA;  31621 MW;  40737F71948B595A CRC64;

  Query Match             61.5%;  Score 67;  DB 1;  Length 314;
  Best Local Similarity   78.3%;  Pred. No. 1;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              ||||| ||||| |||||   |||
Db        244 AAAAAAAAAAAAAAAAAGGLAAA 266


RESULT 13
MAZ_HUMAN
ID   MAZ_HUMAN      STANDARD;      PRT;   477 AA.
AC   P56270; Q15703; Q99443;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Myc-associated zinc finger protein (MAZI) (Purine-binding
DE   transcription factor) (Pur-1) (ZF87) (ZIF87).
GN   MAZ.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92366479; PubMed=1502157;
RA   Bossone S.A., Asselin C., Patel A.J., Marcu K.B.;
RT   "MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences
RT   regulating transcriptional initiation and termination.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Carcinoma;
RX   MEDLINE=92232709; PubMed=1567856;
RA   Pyrc J.J., Moberg K.H., Hall D.J.;
RT   "Isolation of a novel cDNA encoding a zinc-finger protein that binds
RT   to two sites within the c-myc promoter.";
RL   Biochemistry 31:4102-4110(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreatic islets;
RX   MEDLINE=96428591; PubMed=8831693;
RA   Tsutsui H., Sakatsume O., Itakura K., Yokoyama K.K.;
RT   "Members of the MAZ family: a novel cDNA clone for MAZ from human
RT   pancreatic islet cells.";
RL   Biochem. Biophys. Res. Commun. 226:801-809(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96224025; PubMed=8626793;
```

```
RA   Parks C.L., Shenk T.;
RT   "The serotonin 1a receptor gene contains a TATA-less promoter that
RT   responds to MAZ and Sp1.";
RL   J. Biol. Chem. 271:4417-4430(1996).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymphoblastoma;
RX   MEDLINE=98352105; PubMed=9685418;
RA   Song J., Murakami H., Tsutsui H., Tang X., Matsumura M., Itakura K.,
RA   Kanazawa I., Sun K., Yokoyama K.K.;
RT   "Genomic organization and expression of a human gene for Myc-
RT   associated zinc finger protein (MAZ).";
RL   J. Biol. Chem. 273:20603-20614(1998).
CC   -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES
CC       IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,
CC       ME1A1 AND ME1A2, WITHIN THE C-MYC PROMOTER HAVING GREATER
CC       AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES
CC       WITHIN THE PROMOTER OF THE SP1 FAMILY OF TRANSCRIPTION FACTORS.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL
CC       MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.
CC   -!- SIMILARITY: Contains 6 C2H2-type zinc fingers.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M94046; -; NOT_ANNOTATED_CDS.
DR   EMBL; M93339; -; NOT_ANNOTATED_CDS.
DR   EMBL; D85131; BAA12728.1; ALT_INIT.
DR   EMBL; U33819; AAB04121.1; ALT_INIT.
DR   EMBL; AB017335; BAA33064.1; -.
DR   PIR; A42170; A42170.
DR   TRANSFAC; T00490; -.
DR   TRANSFAC; T02305; -.
DR   Genew; HGNC:6914; MAZ.
DR   MIM; 600999; -.
DR   GO; GO:0006367; P:transcription initiation from Pol II promoter; TAS.
DR   GO; GO:0006369; P:transcription termination from Pol II promoter; TAS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 6.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 6.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   RNA-binding; Repeat; Nuclear protein.
FT   ZN_FING     190    212       C2H2-TYPE 1.
FT   ZN_FING     279    301       C2H2-TYPE 2.
FT   ZN_FING     307    329       C2H2-TYPE 3.
FT   ZN_FING     337    360       C2H2-TYPE 4.
FT   ZN_FING     366    388       C2H2-TYPE 5.
FT   ZN_FING     392    413       C2H2-TYPE 6.
FT   DOMAIN       96    108       POLY-ALA.
FT   DOMAIN      133    139       POLY-PRO.
FT   DOMAIN      157    161       POLY-ALA.
FT   DOMAIN      245    249       POLY-GLY.
FT   DOMAIN      435    449       POLY-ALA.
FT   CONFLICT    259    259       MISSING (IN REF. 3).
FT   CONFLICT    401    401       L -> M (IN REF. 2 AND 4).
FT   CONFLICT    443    447       MISSING (IN REF. 3).
SQ   SEQUENCE   477 AA;  48607 MW;  C04C80F32C3C6825 CRC64;

  Query Match             61.5%;  Score 67;  DB 1;  Length 477;
  Best Local Similarity   77.3%;  Pred. No. 1.4;
  Matches   17;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 AAAAEAAAAAEAAAAAEAAAAA 25
              ||| |:|||| ||||| || ||
Db         90 AAAQESAAAAAAAAAAAAAVAA 111


RESULT 14
HM1D_DROAN
ID   HM1D_DROAN     STANDARD;      PRT;   606 AA.
AC   P22544;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Homeobox protein B-H1.
GN   B-H1 OR OM(1D).
OS   Drosophila ananassae (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7217;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91122048; PubMed=1671353;
RA   Tanda S., Corces V.G.;
RT   "Retrotransposon-induced overexpression of a homeobox gene causes
RT   defects in eye morphogenesis in Drosophila.";
RL   EMBO J. 10:407-417(1991).
CC   -!- FUNCTION: Probably involved in eye morphogenesis.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE ANTP HOMEOBOX FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X56682; CAA40011.1; -.
DR   PIR; S13367; S13367.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T03732; -.
DR   FlyBase; FBgn0012114; Dana\B-H1.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   PRINTS; PR00024; HOMEOBOX.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Homeobox; Developmental protein; Nuclear protein; Vision.
FT   DOMAIN       23     57       HIS/GLN-RICH (OPA-REPEAT).
FT   DOMAIN      106    124       HIS/GLN-RICH (OPA-REPEAT).
FT   DOMAIN      173    193       HIS/PRO-RICH.
FT   DNA_BIND    331    390       HOMEOBOX.
FT   DOMAIN      220    248       ALA-RICH.
FT   DOMAIN      422    434       ALA-RICH.
FT   DOMAIN      450    455       ALA-RICH.
FT   DOMAIN      503    510       ALA-RICH.
FT   DOMAIN      515    521       PRO-RICH.
SQ   SEQUENCE   606 AA;  61735 MW;  AA7B8B6367370FBB CRC64;

  Query Match             61.5%;  Score 67;  DB 1;  Length 606;
  Best Local Similarity   78.3%;  Pred. No. 1.7;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              ||||| |||    ||||| |||||
Db        220 AAAAAAAAARNSAAAAAAAAAAA 242


RESULT 15
HXAD_MOUSE
ID   HXAD_MOUSE     STANDARD;      PRT;   386 AA.
AC   Q62424;
```

```
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Homeobox protein Hox-A13 (Hox-1.10).
GN   HOXA13 OR HOX-1.10.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96259555; PubMed=8673126;
RA   Mortlock D.P., Post L.C., Innis J.W.;
RT   "The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads
RT   to arrest of digital arch formation.";
RL   Nat. Genet. 13:284-289(1996).
CC   -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC       A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC       SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: DEFECTS IN HOXA13 ARE THE CAUSE OF HYPODACTYLY (HD), A
CC       CONDITION CHARACTERIZED BY PROOFOUND DEFICIENCY OF DIGITAL ARCH
CC       STRUCTURES.
CC   -!- SIMILARITY: BELONGS TO THE ABD-B HOMEOBOX FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U59322; AAB03322.1; -.
DR   HSSP; P14653; 1B72.
DR   TRANSFAC; T03337; -.
DR   MGD; MGI:96173; Hoxa13.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   ProDom; PD000010; Homeobox; 1.
DR   SMART; SM00389; HOX; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW   Transcription regulation.
FT   DOMAIN       38     51       POLY-ALA.
FT   DNA_BIND    320    379       HOMEOBOX.
FT   DOMAIN       52     57       POLY-GLY.
FT   DOMAIN       62     66       POLY-ALA.
FT   DOMAIN       73     84       POLY-ALA.
FT   DOMAIN      101    104       POLY-ALA.
FT   DOMAIN      116    133       POLY-ALA.
FT   DOMAIN      198    205       POLY-ALA.
SQ   SEQUENCE   386 AA;  39566 MW;  2B01DCC9B1951324 CRC64;

  Query Match             60.6%;  Score 66;  DB 1;  Length 386;
  Best Local Similarity   70.8%;  Pred. No. 1.5;
  Matches   17;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              ||  :| ||||| ||||| ||||:
Db        111 EAPPSAAAAAAAAAAAAAAAAAAS 134

Search completed: January 30, 2004, 00:20:45
Job time : 6.6338 secs
```

```
OM protein - protein search, using sw model

Run on:         January 29, 2004, 23:58:52 ; Search time 25.8216 Seconds
                                             (without alignments)
                                             249.842 Million cell updates/sec

Title:          US-09-461-684C-2
Perfect score:  109
Sequence:       1 CEAAAAAEAAAAAEAAAAAEAAAAA 25

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	84	77.1	1038	5	Q8MQW9	Q8mqw9 drosophila
2	84	77.1	2347	5	Q8INH9	Q8inh9 drosophila
3	84	77.1	2451	5	Q9VG05	Q9vg05 drosophila
4	82	75.2	378	11	Q8R089	Q8r089 mus musculu
5	80	73.4	301	11	Q8BJK2	Q8bjk2 mus musculu
6	80	73.4	1354	11	Q9EPW8	Q9epw8 mus musculu
7	79	72.5	110	11	Q91WW0	Q91ww0 mus musculu
8	79	72.5	332	4	Q8IZU1	Q8izu1 homo sapien
9	77	70.6	265	10	Q39598	Q39598 chlamydomon
10	76	69.7	665	11	Q9QXG2	Q9qxg2 mus musculu
11	76	69.7	1452	4	Q9H4A0	Q9h4a0 homo sapien
12	76	69.7	1512	4	Q9H4A1	Q9h4a1 homo sapien
13	75	68.8	221	5	Q9VXD3	Q9vxd3 drosophila
14	72	66.1	997	5	Q9W2J2	Q9w2j2 drosophila
15	71	65.1	206	10	Q39597	Q39597 chlamydomon
16	71	65.1	1216	10	Q9SWH3	Q9swh3 chlamydomon
17	71	65.1	1787	10	Q9M4X9	Q9m4x9 chlamydomon
18	70	64.2	246	16	Q8G3U2	Q8g3u2 bifidobacte
19	70	64.2	323	4	Q9H782	Q9h782 homo sapien
20	70	64.2	349	4	O43856	O43856 homo sapien
21	70	64.2	423	4	O75400	O75400 homo sapien
22	70	64.2	452	4	O75404	O75404 homo sapien
23	70	64.2	484	11	Q923T4	Q923t4 mus musculu
24	70	64.2	512	4	Q8NAP8	Q8nap8 homo sapien
25	70	64.2	618	11	Q8BX04	Q8bx04 mus musculu
26	70	64.2	618	11	Q8BUW9	Q8buw9 mus musculu
27	70	64.2	1171	3	Q9P3E2	Q9p3e2 neurospora
28	69	63.3	218	11	Q9CS14	Q9cs14 mus musculu
29	69	63.3	512	11	Q8CBI2	Q8cbi2 mus musculu
30	69	63.3	1222	5	Q8T8L9	Q8t8l9 drosophila
31	69	63.3	1354	5	Q8MPN4	Q8mpn4 drosophila
32	69	63.3	1668	5	O76930	O76930 drosophila
33	69	63.3	1669	5	Q9V727	Q9v727 drosophila
34	68	62.4	403	16	Q9JUT2	Q9jut2 neisseria m
35	68	62.4	543	5	Q9W4F9	Q9w4f9 drosophila
36	68	62.4	545	12	Q91TR2	Q91tr2 tupaia herp
37	68	62.4	1340	16	Q9L1H8	Q9l1h8 streptomyce
38	68	62.4	2023	5	Q9V5Z9	Q9v5z9 drosophila
39	68	62.4	2023	5	O96542	O96542 drosophila
40	67.5	61.9	324	5	Q9N6K2	Q9n6k2 drosophila
41	67.5	61.9	324	5	Q9NGI7	Q9ngi7 drosophila
42	67.5	61.9	324	5	Q9NGI8	Q9ngi8 drosophila
43	67.5	61.9	324	5	Q9N6K3	Q9n6k3 drosophila
44	67.5	61.9	324	5	Q9NGI9	Q9ngi9 drosophila
45	67	61.5	324	5	Q9NGB4	Q9ngb4 drosophila

ALIGNMENTS

```
RESULT 1
Q8MQW9
ID   Q8MQW9     PRELIMINARY;     PRT;  1038 AA.
AC   Q8MQW9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   SD05989p (Fragment).
GN   CG7518.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY122252; AAM52764.1; -.
DR   FlyBase; FBgn0038108; CG7518.
DR   InterPro; IPR000104; Antifreeze_1.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER      1      1
SQ   SEQUENCE   1038 AA;  109059 MW;  80C935A2C6D8A276 CRC64;

  Query Match             77.1%;  Score 84;  DB 5;  Length 1038;
  Best Local Similarity   91.3%;  Pred. No. 1.1;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              |||||||||||||||| | |||
Db        112 AAAAAEAAAAAEAAAAAAAEAAA 134
```

```
RESULT 2
Q8INH9
ID   Q8INH9      PRELIMINARY;      PRT;  2347 AA.
AC   Q8INH9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG7518-PB.
GN   CG7518.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003698; AAN14338.1; -.
SQ   SEQUENCE   2347 AA;  257013 MW;  23BF5FC5FFCAEA64 CRC64;

  Query Match             77.1%;  Score 84;  DB 5;  Length 2347;
  Best Local Similarity   91.3%;  Pred. No. 2.3;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              |||||||||||||||||| | |||
Db       1374 AAAAAEAAAAAEAAAAAAAEAAA 1396


RESULT 3
Q9VG05
ID   Q9VG05      PRELIMINARY;      PRT;  2451 AA.
AC   Q9VG05;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG7518 protein.
GN   CG7518.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003698; AAF54888.2; -.
DR   FlyBase; FBgn0038108; CG7518.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   PROSITE; PS00037; MYB_1; 1.
SQ   SEQUENCE   2451 AA;  266959 MW;  088A2293F27481E2 CRC64;

  Query Match            77.1%;  Score 84;  DB 5;  Length 2451;
  Best Local Similarity  91.3%;  Pred. No. 2.4;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAAEAAAAA 25
              |||||||||||||||||| | |||
Db       1374 AAAAAEAAAAAEAAAAAAAAEAAA 1396


RESULT 4
Q8R089
ID   Q8R089      PRELIMINARY;      PRT;   378 AA.
AC   Q8R089;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to hypothetical protein FLJ11618.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC027193; AAH27193.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   378 AA;  39456 MW;  4C3FAF0D4AC29E69 CRC64;

  Query Match            75.2%;  Score 82;  DB 11;  Length 378;
  Best Local Similarity  87.5%;  Pred. No. 0.73;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAAEAAAAA 25
              |||||| ||||| ||||| |||||
Db        222 EAAAAAAAAAAAAAAAAAAAAAAA 245


RESULT 5
Q8BJK2
ID   Q8BJK2      PRELIMINARY;      PRT;   301 AA.
AC   Q8BJK2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cell division cycle 2-like 5 (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK083577; BAC38957.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   301 AA;  32269 MW;  A6CF891DBE25E09E CRC64;

  Query Match            73.4%;  Score 80;  DB 11;  Length 301;
  Best Local Similarity  83.3%;  Pred. No. 0.93;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAAEAAAAA 25
              ||| ||||| ||||| |||||| |
Db         74 EAARAAEAAKAAEAAKAAEAAAKA 97


RESULT 6
Q9EPW8
ID   Q9EPW8      PRELIMINARY;      PRT;  1354 AA.
AC   Q9EPW8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Nischarin.
GN   NISCH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR outbred; TISSUE=Brain;
RX   MEDLINE=20571837; PubMed=11121431;
RA   Alahari S.K., Lee J.W., Juliano R.L.;
RT   "Nischarin, a Novel Protein That Interacts with the Integrin alpha5
RT   Subunit and Inhibits Cell Migration.";
RL   J. Cell Biol. 151:1141-1154(2000).
DR   EMBL; AF315344; AAG42100.1; -.
DR   MGD; MGI:1928323; Nisch.
DR   InterPro; IPR001128; Cytochrome_P450.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007092; LRR_SDS22.
DR   Pfam; PF00560; LRR; 5.
DR   PRINTS; PR00019; LEURICHRPT.
DR   PROSITE; PS00086; CYTOCHROME_P450; 1.
DR   PROSITE; PS50504; LRR_SDS22; 1.
SQ   SEQUENCE   1354 AA;  148060 MW;  01BD676FDCA19247 CRC64;

  Query Match            73.4%;  Score 80;  DB 11;  Length 1354;
  Best Local Similarity  83.3%;  Pred. No. 3.4;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAAEAAAAA 25
              || ||||| ||||| ||||| |||
Db        837 EAPAAAEAPAAAEAPAAAEAPAAA 860


RESULT 7
Q91WW0
ID   Q91WW0      PRELIMINARY;      PRT;   110 AA.
AC   Q91WW0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 10.7 kDa protein.
GN   AI591529.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Salivary gland;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC012681; AAH12681.1; -.
DR   MGD; MGI:2139008; AI591529.
DR   InterPro; IPR000104; Antifreeze_1.
DR   InterPro; IPR001859; Ribosomal_P2.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00456; RIBOSOMALP2.
KW   Hypothetical protein.
SQ   SEQUENCE   110 AA;  10662 MW;  0581D2F635F87EAB CRC64;

  Query Match             72.5%;  Score 79;  DB 11;  Length 110;
  Best Local Similarity   83.3%;  Pred. No. 0.49;
  Matches   20;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              :||||| ||||| ||||| |||||
Db         23 DAAAAAAAAAAAAAAAAAAAAAAA 46


RESULT 8
Q8IZU1
ID   Q8IZU1      PRELIMINARY;      PRT;   332 AA.
AC   Q8IZU1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Family with sequence similarity 9, member A.
GN   FAM9A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22202142; PubMed=12213195;
RA   Martinez-Garay I., Jablonka S., Sutajova M., Steuernagel P., Gal A.,
RA   Kutsche K.;
RT   "A New Gene Family (FAM9) of Low-Copy Repeats in Xp22.3 Expressed
RT   Exclusively in Testis: Implications for Recombinations in This
RT   Region.";
RL   Genomics 80:259-267(2002).
DR   EMBL; AF494343; AAN07162.1; -.
SQ   SEQUENCE   332 AA;  37339 MW;  92F22EC36038229C CRC64;

  Query Match             72.5%;  Score 79;  DB 4;  Length 332;
  Best Local Similarity   83.3%;  Pred. No. 1.3;
  Matches   20;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              ||  |   ||||||||||||||||
Db        190 EAEEAEAAAAAAEAAAAAEAAAAA 213


RESULT 9
Q39598
ID   Q39598      PRELIMINARY;      PRT;   265 AA.
AC   Q39598;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cgcr-4 product (Fragment).
GN   CGCR-4.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92119224; PubMed=1731966;
RA   Wakarchuk W.W., Muller F.W., Beck C.F.;
RT   "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
RT   arrangements of directly repeated sequence motifs.";
RL   Plant Mol. Biol. 18:143-146(1992).
DR   EMBL; X17208; CAA35080.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   265 AA;  26216 MW;  B35318B7377CF782 CRC64;

  Query Match             70.6%;  Score 77;  DB 10;  Length 265;
  Best Local Similarity   82.6%;  Pred. No. 1.6;
  Matches   19;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAA 25
              |||||||||||:|  |||| |||
Db        154 AAAAAEAAAAAKARVAAEARAAA 176


RESULT 10
Q9QXG2
ID   Q9QXG2      PRELIMINARY;      PRT;   665 AA.
AC   Q9QXG2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Choroideremia protein.
GN   CHM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   van den Hurk J.A., Huber I., van de Pol T.J., Cremers F.P.;
RT   "Cloning and sequencing of the mouse choroideremia gene.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF218084; AAF25478.1; -.
DR   MGD; MGI:892979; Chm.
DR   InterPro; IPR002005; Rab_GDI_REP.
DR   Pfam; PF00996; GDI; 1.
DR   PRINTS; PR00891; RABGDIREP.
SQ   SEQUENCE   665 AA;  73976 MW;  FF71A74AD3FBDE0A CRC64;

  Query Match             69.7%;  Score 76;  DB 11;  Length 665;
  Best Local Similarity   79.2%;  Pred. No. 4.4;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              ||| ||||| | ||| ||||| ||
Db        132 EAAEAAEAAEATEAAEAAEAAEAA 155


RESULT 11
Q9H4A0
ID   Q9H4A0      PRELIMINARY;      PRT;  1452 AA.
AC   Q9H4A0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CDC2L5 protein kinase.
GN   CDC2L5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,
RA   Picard A., Callebaut I., Perre E., Geneviere A.M.;
RT   "A new subfamily of high molecular mass CDC2-related kinases with
```

```
RT   PITAI/VRE.";
RL   Biochem. Biophys. Res. Commun. 279:832-837(2001).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AJ297710; CAC10401.1; -.
DR   HSSP; P24941; 1BUH.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   1452 AA;  158480 MW;  C7ED072968B439CB CRC64;

  Query Match             69.7%;  Score 76;  DB 4;  Length 1452;
  Best Local Similarity   79.2%;  Pred. No. 8.6;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              ||| ||||| ||||  |||||| |
Db        463 EAARAAEAAKAAEATKAAEAAAKA 486


RESULT 12
Q9H4A1
ID   Q9H4A1      PRELIMINARY;      PRT;  1512 AA.
AC   Q9H4A1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CDC2L5 protein kinase.
GN   CDC2L5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Marques F., Moreau J.L., Peaucellier G., Lozano J.C., Schatt P.,
RA   Picard A., Callebaut I., Perre E., Geneviere A.M.;
RT   "A new subfamily of high molecular mass CDC2-related kinases with
RT   PITAI/VRE.";
RL   Biochem. Biophys. Res. Commun. 279:832-837(2001).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; AJ297709; CAC10400.1; -.
DR   HSSP; P24941; 1BUH.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   1512 AA;  164969 MW;  283B8D553DB57650 CRC64;

  Query Match             69.7%;  Score 76;  DB 4;  Length 1512;
  Best Local Similarity   79.2%;  Pred. No. 8.9;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              ||| ||||| ||||  |||||| |
Db        463 EAARAAEAAKAAEATKAAEAAAKA 486


RESULT 13
Q9VXD3
ID   Q9VXD3      PRELIMINARY;      PRT;   221 AA.
AC   Q9VXD3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CG13012 protein.
GN   CG13012.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003503; AAF48641.1; -.
DR   FlyBase; FBgn0030769; CG13012.
SQ   SEQUENCE   221 AA;  22987 MW;  A1B95919B167C5E2 CRC64;

  Query Match             68.8%;  Score 75;  DB 5;  Length 221;
  Best Local Similarity   81.8%;  Pred. No. 2.1;
  Matches   18;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          4 AAAAEAAAAAEAAAAAEAAAAA 25
              |||||||| ||||  ||||| |
Db          6 AAAAEAAAVAEAATVAEAAAVA 27


RESULT 14
Q9W2J2
ID   Q9W2J2      PRELIMINARY;      PRT;   997 AA.
AC   Q9W2J2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG18375 protein.
```

```
GN   CG18375.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003453; AAF46699.2; -.
DR   HSSP; Q13625; 1YCS.
DR   FlyBase; FBgn0034606; CG18375.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000104; Antifreeze_1.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00023; ank; 2.
DR   PRINTS; PR01415; ANKYRIN.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00248; ANK; 2.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   ANK repeat; Repeat.
SQ   SEQUENCE   997 AA;  107821 MW;  E712D400C2C4FD3D CRC64;

  Query Match             66.1%;  Score 72;  DB 5;  Length 997;
  Best Local Similarity   79.2%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              |||||| ||||| ||||  || ||
Db        432 EAAAAAAAAAAAAAAAAQAAAEAA 455


RESULT 15
Q39597
ID   Q39597     PRELIMINARY;      PRT;   206 AA.
AC   Q39597;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cgcr-1 product (Fragment).
GN   CGCR-1.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cw15 mt-;
RX   MEDLINE=92119224; PubMed=1731966;
RA   Wakarchuk W.W., Muller F.W., Beck C.F.;
RT   "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
RT   arrangements of directly repeated sequence motifs.";
RL   Plant Mol. Biol. 18:143-146(1992).
DR   EMBL; X17207; CAA35079.1; -.
FT   NON_TER       1      1
FT   NON_TER     206    206
SQ   SEQUENCE   206 AA;  19869 MW;  ED3FF120EF8EFAE1 CRC64;

  Query Match             65.1%;  Score 71;  DB 10;  Length 206;
  Best Local Similarity   75.0%;  Pred. No. 4.7;
  Matches   18;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAA 25
              | ||||:||||||||||   || |
Db         49 EEAAAAQAAAAAEAAAAAEARAAPA 72
```

Search completed: January 30, 2004, 00:24:39
Job time : 27.8216 secs

```
OM protein - protein search, using sw model

Run on:         January 30, 2004, 00:13:12 ; Search time 10.2535 Seconds
                                             (without alignments)
                                             225.098 Million cell updates/sec

Title:          US-09-461-684C-3
Perfect score:  143
Sequence:       1 CGLFGAIAGFIENGWEGMIDGWYG 24

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_76:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	134	93.7	550	1	HMIVS2	hemagglutinin prec
2	134	93.7	550	1	HMIVS3	hemagglutinin prec
3	134	93.7	550	1	HMIV77	hemagglutinin prec
4	134	93.7	550	1	HMIV80	hemagglutinin prec
5	134	93.7	550	1	HMIV33	hemagglutinin prec
6	134	93.7	550	1	HMIV89	hemagglutinin prec
7	134	93.7	550	1	HMIV21	hemagglutinin prec
8	134	93.7	550	1	HMIV98	hemagglutinin prec
9	134	93.7	550	1	HMIV15	hemagglutinin prec
10	134	93.7	550	2	JQ1153	hemagglutinin prec
11	134	93.7	550	2	JQ1154	hemagglutinin prec
12	134	93.7	550	2	JQ1155	hemagglutinin prec
13	134	93.7	566	1	HMIVH	hemagglutinin prec
14	134	93.7	566	1	HMIVHA	hemagglutinin prec
15	134	93.7	566	1	HMIVHM	hemagglutinin prec
16	134	93.7	566	1	HMIVDU	hemagglutinin prec
17	133	93.0	561	1	HMIV49	hemagglutinin prec
18	133	93.0	561	1	HMIV84	hemagglutinin prec
19	132	92.3	565	1	HMIVE1	hemagglutinin prec
20	132	92.3	565	1	HMIVE3	hemagglutinin prec
21	132	92.3	566	1	HMIV6	hemagglutinin prec
22	132	92.3	567	1	HMIVV	hemagglutinin prec
23	131	91.6	362	2	S38637	hemagglutinin - in
24	131	91.6	550	1	HMIV86	hemagglutinin prec
25	131	91.6	560	1	HMIVT7	hemagglutinin prec
26	131	91.6	565	1	HMIVE2	hemagglutinin prec
27	131	91.6	565	1	HMIVE4	hemagglutinin prec
28	131	91.6	565	1	HMIVE5	hemagglutinin prec
29	131	91.6	565	1	HMIVE6	hemagglutinin prec
30	131	91.6	565	1	HMIVE7	hemagglutinin prec
31	131	91.6	565	1	HMIVE8	hemagglutinin prec
32	131	91.6	565	1	HMIVE9	hemagglutinin prec
33	131	91.6	565	1	HMIVET	hemagglutinin prec
34	131	91.6	565	1	HMIVEE	hemagglutinin prec
35	131	91.6	565	2	S33703	hemagglutinin - in
36	131	91.6	570	1	A45591	hemagglutinin prec
37	131	91.6	570	2	S22013	hemagglutinin prec
38	131	91.6	570	2	S22014	hemagglutinin prec
39	131	91.6	570	2	S22015	hemagglutinin prec
40	131	91.6	570	2	S22016	hemagglutinin prec
41	131	91.6	570	2	S22017	hemagglutinin prec
42	131	91.6	570	2	S22018	hemagglutinin prec
43	131	91.6	570	2	S22020	hemagglutinin prec
44	131	91.6	570	2	S22021	hemagglutinin prec
45	131	91.6	570	2	S22029	hemagglutinin prec

ALIGNMENTS

```
RESULT 1
HMIVS2
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C;Species: influenza A virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C;Accession: A29971
R;Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A;Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A;Reference number: A94370; MUID:88101364; PMID:3336940
A;Accession: A29971
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M19056; NID:g324208
A;Note: the sequence in GenBank entry FLAHAPA, release 106, (PID:g324209) differs from tl
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 2
HMIVS3
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C;Species: influenza A virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C;Accession: B29971
R;Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A;Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A;Reference number: A94370; MUID:88101364; PMID:3336940
A;Accession: B29971
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M19057; NID:g324210
A;Note: the sequence in GenBank entry FLAHAPB, release 106, (PID:g324211) differs from tl
C;Genetics:
A;Map position: segment 4
```

```
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 3
HMIV77
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: A27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: A27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16737; NID:g324081; PIDN:AAA43143.1; PID:g324082
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 4
HMIV80
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: B27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: B27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16738; NID:g324083
A;Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 5
HMIV33
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: C27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: C27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 6
HMIV89
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: D27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: D27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
```

```
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 7
HMIV21
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: E27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: E27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16741; NID:g324089
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 8
HMIV98
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: F27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: F27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16742; NID:g324091
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 9
HMIV15
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: G27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: G27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16743; NID:g324093; PIDN:AAA43149.1; PID:g324094
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 10
JQ1153
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1153
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1153
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00929; NID:g221279; PIDN:BAA00769.1; PID:g221280
A;Note: the authors translated the codon GGG for residue 218 as Glu
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 2;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
```

```
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 11
JQ1154
hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1154
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1154
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00930; NID:g221273; PIDN:BAA00770.1; PID:g221274
A;Note: the authors translated the codon GGG for residue 218 as Glu
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 2;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 12
JQ1155
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1155
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1155
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00931; NID:g221277; PIDN:BAA00771.1; PID:g221278
A;Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 538
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 2;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 13
HMIVH
hemagglutinin precursor - influenza A virus
C;Species: influenza A virus
C;Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999
C;Accession: A93705; A93233; A04051; A93231; A94441
R;Both, G.W.; Sleigh, M.J.
Nucleic Acids Res. 8, 2561-2575, 1980
A;Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza v
A;Reference number: A93705; MUID:81053698; PMID:6253883
A;Accession: A93705
A;Molecule type: genomic RNA
A;Residues: 1-566 <BOT>
A;Cross-references: GB:V01103
A;Experimental source: strain A/NT/60/68/29C
A;Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/6
R;Dopheide, T.A.; Ward, C.W.
FEBS Lett. 110, 181-183, 1980
A;Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.
A;Reference number: A91276; MUID:80179105; PMID:6768586
A;Contents: annotation; disulfide bonds
R;Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.
Nature 287, 301-306, 1980
A;Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from
A;Reference number: A93233; MUID:81030852; PMID:7421990
A;Accession: A93233
A;Molecule type: genomic RNA
A;Residues: 1-24,'S',26,'D',28-159,'G',161-197,'I',199-241,'L',243-249 <GET>
A;Experimental source: strain X-31[H3]
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;536-552/Domain: transmembrane #status predicted <TM1>
F;30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 14
HMIVHA
hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C;Accession: A93231; A04051
R;Verhoeyen, M.; Fang, R.; Min Jou, W.; Devos, R.; Huylebroeck, D.; Saman, E.; Fiers, W.
Nature 286, 771-776, 1980
A;Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/
A;Reference number: A93231; MUID:80254693; PMID:7402351
A;Accession: A93231
A;Molecule type: genomic RNA
A;Residues: 1-566 <VER>
A;Cross-references: GB:J02090; NID:g324131; PIDN:AAA43178.1; PID:g324132
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 15
```

```
HMIVHM
hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 31-Mar-2000
C;Accession: A94441; A04051
R;Sleigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.
in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, E
A;Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of c
A;Reference number: A94441
A;Accession: A94441
A;Molecule type: genomic RNA
A;Residues: 1-566 <SLE>
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             93.7%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


Search completed: January 30, 2004, 00:26:21
Job time : 10.2535 secs
```

```
OM protein - protein search, using sw model

Run on:         January 29, 2004, 23:58:27 ; Search time 5.40845 Seconds
                                             (without alignments)
                                             208.681 Million cell updates/sec

Title:          US-09-461-684C-3
Perfect score:  143
Sequence:       1 CGLFGAIAGFIENGWEGMIDGWYG 24

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	134	93.7	550	1	HEMA_IADH1	P12582 influenza a
2	134	93.7	550	1	HEMA_IADH2	P12583 influenza a
3	134	93.7	550	1	HEMA_IADH3	P12584 influenza a
4	134	93.7	550	1	HEMA_IADH4	P12585 influenza a
5	134	93.7	550	1	HEMA_IADH5	P12586 influenza a
6	134	93.7	550	1	HEMA_IADH6	P12587 influenza a
7	134	93.7	550	1	HEMA_IADH7	P12588 influenza a
8	134	93.7	550	1	HEMA_IADHK	P43257 influenza a
9	134	93.7	550	1	HEMA_IADHL	P43258 influenza a
10	134	93.7	550	1	HEMA_IAGHK	P43260 influenza a
11	134	93.7	550	1	HEMA_IAZH2	P11133 influenza a
12	134	93.7	550	1	HEMA_IAZH3	P11134 influenza a
13	134	93.7	566	1	HEMA_IAAIC	P03437 influenza a
14	134	93.7	566	1	HEMA_IADA3	P26134 influenza a
15	134	93.7	566	1	HEMA_IADU3	P03442 influenza a
16	134	93.7	566	1	HEMA_IAMAO	P26138 influenza a
17	134	93.7	566	1	HEMA_IAME1	P03449 influenza a
18	134	93.7	566	1	HEMA_IAME2	P03439 influenza a
19	134	93.7	566	1	HEMA_IANT6	P03436 influenza a
20	133	93.0	561	1	HEMA_IACKG	P12581 influenza a
21	133	93.0	561	1	HEMA_IAMIN	P12439 influenza a
22	132	92.3	565	1	HEMA_IAHTO	P17000 influenza a
23	132	92.3	565	1	HEMA_IAHUR	P17002 influenza a
24	132	92.3	566	1	HEMA_IAEN7	P03440 influenza a
25	132	92.3	566	1	HEMA_IAZCO	P26139 influenza a
26	132	92.3	567	1	HEMA_IAVI7	P03435 influenza a
27	131	91.6	550	1	HEMA_IAME6	P12589 influenza a
28	131	91.6	560	1	HEMA_IATKR	P03458 influenza a
29	131	91.6	565	1	HEMA_IAHAL	P16994 influenza a
30	131	91.6	565	1	HEMA_IAHFO	P16995 influenza a
31	131	91.6	565	1	HEMA_IAHK6	P19699 influenza a
32	131	91.6	565	1	HEMA_IAHK7	P16996 influenza a
33	131	91.6	565	1	HEMA_IAHMI	P15658 influenza a
34	131	91.6	565	1	HEMA_IAHNM	P16997 influenza a
35	131	91.6	565	1	HEMA_IAHRO	P16998 influenza a
36	131	91.6	565	1	HEMA_IAHSA	P16999 influenza a
37	131	91.6	565	1	HEMA_IAHSU	Q08011 influenza a
38	131	91.6	565	1	HEMA_IAHTE	P17001 influenza a
39	131	91.6	566	1	HEMA_IAZUK	P26141 influenza a
40	131	91.6	570	1	HEMA_IAHC6	P26094 influenza a
41	131	91.6	570	1	HEMA_IAHC7	P26095 influenza a
42	131	91.6	570	1	HEMA_IAHCD	P26096 influenza a
43	131	91.6	570	1	HEMA_IAHDE	P26097 influenza a
44	131	91.6	570	1	HEMA_IAHLE	P26098 influenza a
45	131	91.6	570	1	HEMA_IAHLO	P26099 influenza a

ALIGNMENTS

```
RESULT 1
HEMA_IADH1
ID   HEMA_IADH1     STANDARD;      PRT;   550 AA.
AC   P12582; Q84021; Q84022;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/5/77).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11357;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16737; AAA43143.1; -.
DR   HSSP; P03437; 3HMG.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61705 MW;  7E7ACFE716FC969A CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 2
HEMA_IADH2
ID   HEMA_IADH2     STANDARD;      PRT;   550 AA.
AC   P12583; Q84011;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/8/80).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11358;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16738; AAA43144.1; -.
DR   PIR; B27813; HMIV80.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    137    137       K -> N (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61659 MW;  A107023ACC9CC353 CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 3
HEMA_IADH3
ID   HEMA_IADH3     STANDARD;      PRT;   550 AA.
AC   P12584; Q84012; Q89793;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11359;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16739; AAA43145.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61577 MW;  6C30BF67CFDCB7DE CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 4
HEMA_IADH4
ID   HEMA_IADH4     STANDARD;      PRT;   550 AA.
AC   P12585; Q84013; Q84014;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11360;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16740; AAA43146.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61664 MW;  A16B2CF8CBBBD9D0 CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 5
HEMA_IADH5
ID   HEMA_IADH5     STANDARD;      PRT;   550 AA.
AC   P12586; Q84015; Q84016;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11361;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16741; AAA43147.1; -.
DR   PIR; E27813; HMIV21.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      7      7       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    178    179       YV -> VI (IN PIR DATA BANK).
FT   CONFLICT    388    388       K -> T (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61856 MW;  48401C867A15BF8C CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 6
HEMA_IADH6
ID   HEMA_IADH6     STANDARD;      PRT;   550 AA.
AC   P12587; Q84017;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11362;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16742; AAA43148.1; -.
DR   PIR; F27813; HMIV98.
DR   HSSP; P03437; 1HGJ.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      8      8       Y -> N (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61711 MW;  67BCD85F44736CFE CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 7
HEMA_IADH7
ID   HEMA_IADH7     STANDARD;      PRT;   550 AA.
AC   P12588; Q84018; Q89470;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11363;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16743; AAA43149.1; -.
DR   HSSP; P03437; 3HMG.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61761 MW;  6EF81793281D53EB CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 8
HEMA_IADHK
ID   HEMA_IADHK     STANDARD;      PRT;   550 AA.
AC   P43257;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hong Kong/7/75).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11364;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D00929; BAA00769.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61549 MW;  864639B829FE1BA9 CRC64;
```

```
  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 9
HEMA_IADHL
ID   HEMA_IADHL     STANDARD;      PRT;   550 AA.
AC   P43258;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hong Kong/64/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=45412;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D00931; BAA00771.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61718 MW;  A351C56789E4BE9A CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 10
HEMA_IAGHK
ID   HEMA_IAGHK     STANDARD;      PRT;   550 AA.
AC   P43260;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=45414;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D00930; BAA00770.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61676 MW;  9A1E094DA28BACD2 CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 11
HEMA_IAZH2
ID   HEMA_IAZH2     STANDARD;      PRT;   550 AA.
AC   P11133; Q84019; Q84020;
DT   01-JUL-1989 (Rel. 11, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update).
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE   chain] (Fragment).
GN   HA.
```

```
OS   Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11497;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88101364; PubMed=3336940;
RA   Kida H., Shortridge K.F., Webster R.G.;
RT   "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
RT   in China.";
RL   Virology 162:160-166(1988).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M19057; AAA43212.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Hemagglutinin; Envelope protein; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61437 MW;  1F2A7E758C531CE8 CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 12
HEMA_IAZH3
ID   HEMA_IAZH3     STANDARD;      PRT;   550 AA.
AC   P11134; Q84025; Q84026;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE   chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11498;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88101364; PubMed=3336940;
RA   Kida H., Shortridge K.F., Webster R.G.;
RT   "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
RT   in China.";
RL   Virology 162:160-166(1988).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M19056; AAA43211.1; ALT_TERM.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Hemagglutinin; Envelope protein; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61580 MW;  991F6D8BC02F24F2 CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 13
HEMA_IAAIC
ID   HEMA_IAAIC     STANDARD;      PRT;   566 AA.
AC   P03437;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Aichi/2/68).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=150147;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=80254693; PubMed=7402351;
RA   Verhoeyen M., Fang R., Min Jou W., Devos R., Huylebroeck D.,
RA   Saman E., Fiers W.;
RT   "Antigenic drift between the haemagglutinin of the Hong Kong
RT   influenza strains A/Aichi/2/68 and A/Victoria/3/75.";
RL   Nature 286:771-776(1980).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX   MEDLINE=81123029; PubMed=7464906;
RA   Wilson I.A., Skehel J.J., Wiley D.C.;
RT   "Structure of the haemagglutinin membrane glycoprotein of influenza
RT   virus at 3-A resolution.";
RL   Nature 289:366-373(1981).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE=88232903; PubMed=3374584;
```

```
RA   Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J.,
RA   Wiley D.C.;
RT   "Structure of the influenza virus haemagglutinin complexed with its
RT   receptor, sialic acid.";
RL   Nature 333:426-431(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
RX   MEDLINE=90107940; PubMed=2295311;
RA   Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J.,
RA   Wiley D.C.;
RT   "The structure of a membrane fusion mutant of the influenza virus
RT   haemagglutinin.";
RL   EMBO J. 9:17-24(1990).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE=90230310; PubMed=2329580;
RA   Weis W.I., Bruenger A.T., Skehel J.J., Wiley D.C.;
RT   "Refinement of the influenza virus hemagglutinin by simulated
RT   annealing.";
RL   J. Mol. Biol. 212:737-761(1990).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE=94352388; PubMed=8072525;
RA   Bullough P.A., Hughson F.M., Skehel J.J., Wiley D.C.;
RT   "Structure of influenza haemagglutinin at the pH of membrane fusion.";
RL   Nature 371:37-43(1994).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).
RX   MEDLINE=98120975; PubMed=9461077;
RA   Fleury D., Wharton S.A., Skehel J.J., Knossow M., Bizebard T.;
RT   "Antigen distortion allows influenza virus to escape neutralization.";
RL   Nat. Struct. Biol. 5:119-123(1998).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J02090; AAA43178.1; -.
DR   EMBL; V01085; CAA24269.1; -.
DR   PIR; A93231; HMIVHA.
DR   PDB; 2HMG; 31-OCT-93.
DR   PDB; 3HMG; 31-OCT-93.
DR   PDB; 4HMG; 31-OCT-93.
DR   PDB; 5HMG; 31-JAN-94.
DR   PDB; 1HGD; 31-JAN-94.
DR   PDB; 1HGE; 31-JAN-94.
DR   PDB; 1HGF; 31-JAN-94.
DR   PDB; 1HGG; 31-JAN-94.
DR   PDB; 1HGH; 31-JAN-94.
DR   PDB; 1HGI; 31-JAN-94.
DR   PDB; 1HGJ; 31-JAN-94.
DR   PDB; 1HTM; 14-FEB-95.
DR   PDB; 2VIR; 29-APR-98.
DR   PDB; 2VIS; 29-APR-98.
DR   PDB; 2VIT; 29-APR-98.
DR   PDB; 2VIU; 29-APR-98.
DR   PDB; 1EO8; 29-NOV-00.
DR   PDB; 1HA0; 22-DEC-99.
DR   PDB; 1J8H; 13-MAR-02.
DR   PDB; 1KEN; 24-APR-02.
DR   PDB; 1QFU; 29-DEC-99.
DR   PDB; 1QU1; 05-JAN-00.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   DISULFID     30    482       INTERCHAIN.
FT   DISULFID     68    293
FT   DISULFID     80     92
FT   DISULFID    113    155
FT   DISULFID    297    321
FT   DISULFID    489    493
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .).
FT   STRAND       27     35
FT   STRAND       40     42
FT   STRAND       50     52
FT   STRAND       55     57
FT   STRAND       59     60
FT   STRAND       67     70
FT   STRAND       74     76
FT   TURN         78     79
FT   HELIX        82     87
FT   TURN         88     88
FT   HELIX        90     95
FT   TURN         96     97
FT   STRAND       99     99
FT   STRAND      102    105
FT   TURN        107    108
FT   STRAND      116    117
FT   TURN        119    120
FT   HELIX       121    131
FT   STRAND      133    133
FT   STRAND      136    138
FT   TURN        144    145
FT   STRAND      146    147
FT   STRAND      152    157
FT   TURN        158    159
FT   STRAND      160    162
FT   TURN        165    166
FT   STRAND      167    169
FT   STRAND      171    173
FT   TURN        174    175
FT   STRAND      176    176
FT   STRAND      180    185
FT   STRAND      192    200
FT   HELIX       204    211
FT   STRAND      217    221
FT   STRAND      226    229
FT   STRAND      239    239
FT   TURN        240    241
FT   STRAND      242    242
FT   STRAND      245    253
FT   TURN        255    256
FT   STRAND      258    265
FT   STRAND      267    270
FT   STRAND      272    275
FT   STRAND      282    285
FT   STRAND      290    294
FT   STRAND      297    299
FT   TURN        300    301
FT   STRAND      302    304
FT   STRAND      310    311
FT   STRAND      318    320
FT   STRAND      323    324
FT   STRAND      331    333
FT   STRAND      337    337
FT   TURN        347    348
```

```
FT   TURN        350    350
FT   STRAND      351    351
FT   TURN        352    354
FT   STRAND      355    355
FT   STRAND      359    360
FT   TURN        361    361
FT   STRAND      367    373
FT   TURN        374    375
FT   STRAND      376    382
FT   HELIX       383    400
FT   TURN        401    401
FT   STRAND      406    407
FT   HELIX       421    471
FT   HELIX       472    474
FT   STRAND      475    477
FT   STRAND      482    485
FT   HELIX       491    498
FT   TURN        499    500
FT   HELIX       504    515
FT   TURN        518    519
SQ   SEQUENCE   566 AA;  63415 MW;  E395659C23CAFECA CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 14
HEMA_IADA3
ID   HEMA_IADA3     STANDARD;      PRT;   566 AA.
AC   P26134;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Duck/Alberta/78/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11348;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92114135; PubMed=1731092;
RA   Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
RA   Webster R.G.;
RT   "Evolution of the H3 influenza virus hemagglutinin from human and
RT   nonhuman hosts.";
RL   J. Virol. 66:1129-1138(1992).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M73771; -; NOT_ANNOTATED_CDS.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63534 MW;  FE19AB6FF9415B89 CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 15
HEMA_IADU3
ID   HEMA_IADU3     STANDARD;      PRT;   566 AA.
AC   P03442;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Duck/Ukraine/1/63).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11374;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82025542; PubMed=6169439;
RA   Fang R., Min Jou W., Huylebroeck D., Devos R., Fiers W.;
RT   "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
RT   gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
RT   hemagglutinin.";
RL   Cell 25:315-323(1981).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; V01087; CAA24271.1; -.
DR   PDB; 1IBN; 08-AUG-01.
DR   PDB; 1IBO; 08-AUG-01.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63530 MW;  E70F87F0AE1178F4 CRC64;

  Query Match             93.7%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


Search completed: January 30, 2004, 00:20:45
Job time : 5.40845 secs
```

```
OM protein - protein search, using sw model

Run on:         January 29, 2004, 23:58:52 ; Search time 24.7887 Seconds
                                            (without alignments)
                                            249.842 Million cell updates/sec

Title:          US-09-461-684C-3
Perfect score:  143
Sequence:       1 CGLFGAIAGFIENGWEGMIDGWYG 24

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	134	93.7	384	12	Q8JK63	Q8jk63 influenza a
2	134	93.7	566	12	Q98052	Q98052 influenzavi
3	134	93.7	566	12	Q8UZ51	Q8uz51 influenza a
4	134	93.7	566	12	Q8QLN8	Q8qln8 influenza a
5	134	93.7	566	12	Q67132	Q67132 influenza a
6	134	93.7	566	12	Q67125	Q67125 influenzavi
7	134	93.7	566	12	Q8UXR3	Q8uxr3 influenza a
8	134	93.7	566	12	Q91MA7	Q91ma7 influenza a
9	134	93.7	566	12	Q9DHG0	Q9dhg0 influenza a
10	134	93.7	566	12	Q910M5	Q910m5 influenza a
11	134	93.7	566	12	Q67126	Q67126 influenzavi
12	133	93.0	301	12	Q9DXE3	Q9dxe3 influenza a
13	132	92.3	550	12	Q82499	Q82499 influenzavi
14	132	92.3	550	12	Q82498	Q82498 influenzavi
15	132	92.3	550	12	Q82753	Q82753 influenza v
16	132	92.3	566	12	Q82496	Q82496 influenzavi
17	132	92.3	571	12	Q03909	Q03909 influenza a
18	131	91.6	109	12	Q67050	Q67050 influenzavi
19	131	91.6	109	12	Q67053	Q67053 influenzavi
20	131	91.6	109	12	Q67051	Q67051 influenzavi
21	131	91.6	109	12	Q67052	Q67052 influenzavi
22	131	91.6	362	12	Q9QKD3	Q9qkd3 influenza a
23	131	91.6	362	12	Q9QKD1	Q9qkd1 influenza a
24	131	91.6	362	12	Q82513	Q82513 influenzavi
25	131	91.6	362	12	Q9QKD2	Q9qkd2 influenza a
26	131	91.6	362	12	Q84174	Q84174 influenzavi
27	131	91.6	362	12	Q82517	Q82517 influenzavi
28	131	91.6	365	12	Q9DL25	Q9dl25 influenza a
29	131	91.6	367	12	Q9DL22	Q9dl22 influenza a
30	131	91.6	368	12	Q9DL29	Q9dl29 influenza a
31	131	91.6	369	12	Q9DL26	Q9dl26 influenza a
32	131	91.6	369	12	P87689	P87689 influenzavi
33	131	91.6	369	12	Q9DL06	Q9dl06 influenza a
34	131	91.6	371	12	Q9DL24	Q9dl24 influenza a
35	131	91.6	371	12	P87685	P87685 influenzavi
36	131	91.6	373	12	Q9DL20	Q9dl20 influenza a
37	131	91.6	374	12	Q9DL21	Q9dl21 influenza a
38	131	91.6	375	12	Q9DL27	Q9dl27 influenza a
39	131	91.6	375	12	Q9DL05	Q9dl05 influenza a
40	131	91.6	376	12	Q9DL30	Q9dl30 influenza a
41	131	91.6	376	12	Q9DL04	Q9dl04 influenza a
42	131	91.6	377	12	Q9E7P6	Q9e7p6 influenza a
43	131	91.6	382	12	Q9DL03	Q9dl03 influenza a
44	131	91.6	408	12	Q9E7P5	Q9e7p5 influenza a
45	131	91.6	409	12	Q9Q0L5	Q9q0l5 influenza a

ALIGNMENTS

```
RESULT 1
Q8JK63
ID   Q8JK63      PRELIMINARY;      PRT;   384 AA.
AC   Q8JK63;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hemagglutinin (Fragment).
GN   H3HA.
OS   Influenza A virus (A/teal/Germany/wv201r/01).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=205472;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/teal/Germany/wv01r/01;
RA   Werner O., Starick E., Mueller T., Muehle R.;
RT   "Characterisation of avian influenza virus isolates from wild birds
RT   from Germany.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ506781; CAD44999.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER     384    384
SQ   SEQUENCE   384 AA;  42076 MW;  459731795CA5CE38 CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
```

```
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 2
Q98052
ID   Q98052      PRELIMINARY;      PRT;   566 AA.
AC   Q98052;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin precursor (Fragment).
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81053698; PubMed=6253883;
RA   Both G.W., Sleigh M.J.;
RT   "Complete nucleotide sequence of the haemagglutinin gene from a human
RT   influenza virus of the Hong Kong subtype.";
RL   Nucleic Acids Res. 8:2561-2575(1980).
RN   [2]
RP   SEQUENCE OF 17-344 FROM N.A.
RX   MEDLINE=81194918; PubMed=6164798;
RA   Sleigh M.J., Both G.W., Underwood P.A., Bender V.J.;
RT   "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT   subtype: Correlation of amino acid changes with alterations in viral
RT   antigenicity.";
RL   J. Virol. 37:845-853(1981).
RN   [3]
RP   SEQUENCE OF 17-566 FROM N.A.
RX   MEDLINE=82033276; PubMed=6169843;
RA   Both G.W., Sleigh M.J.;
RT   "Conservation and variation in the hemagglutinins of Hong Kong subtype
RT   influenza viruses during antigenic drift.";
RL   J. Virol. 39:845-853(1981).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; J02135; AAA43189.1; -.
DR   HSSP; P03437; 1HGE.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    344       POTENTIAL.
FT   CHAIN       346    566       POTENTIAL.
SQ   SEQUENCE   566 AA;  63414 MW;  C447FD465BE4FCF9 CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 3
Q8UZ51
ID   Q8UZ51      PRELIMINARY;      PRT;   566 AA.
AC   Q8UZ51;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Haemagglutinin.
GN   HA.
OS   Influenza A virus (A/pet bird/Hong Kong/1559/99(H3N8)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=183665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/pet bird/Hong Kong/1559/99;
RA   Chin P., Shortridge K.F.;
RT   "Characterisation of avian H3 influenza viruses.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ427304; CAD20336.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63403 MW;  F11C91B6A0183484 CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 4
Q8QLN8
ID   Q8QLN8      PRELIMINARY;      PRT;   566 AA.
AC   Q8QLN8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Haemagglutinin.
GN   HA.
OS   Influenza A virus (A/aquatic bird/Hong Kong/399/99(H3N8)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=183664;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/aquatic bird/Hong Kong/399/99;
RA   Chin P., Shortridge K.F.;
RT   "Characterisation of influenza viruses from wild aquatic birds.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ427297; CAD20322.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63412 MW;  68913C222C97B92E CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 5
```

```
Q67132
ID   Q67132      PRELIMINARY;      PRT;   566 AA.
AC   Q67132;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin.
GN   HA.
OS   Influenza A virus (strain A/Aichi/2/68).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=150147;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Aichi/2/68;
RA   Min J.W., Verhoeyen M., Fang R.-X., Devos R., Huylebroeck D.,
RA   Fiers W.;
RT   "Shift and drift in influenza viruses.";
RL   (In) Carlile M.J., Collins J.F., Moseley B.E. B. (eds.);
RL   SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,
RL   Cambridge University Press, New York (1981).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; M55059; AAA43239.1; -.
DR   HSSP; P03437; 1HGE.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   CHAIN         1    344       HEMAGGLUTININ.
FT   CHAIN       346    566       NEURAMINIDASE.
SQ   SEQUENCE   566 AA;  63441 MW;  E5D1B97DF96FFECA CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 6
Q67125
ID   Q67125      PRELIMINARY;      PRT;   566 AA.
AC   Q67125;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin.
GN   HA.
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Seal/MA/3911/92;
RX   MEDLINE=95146951; PubMed=7844533;
RA   Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT   "The appearance of H3 influenza viruses in seals.";
RL   J. Gen. Virol. 76:199-203(1995).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; L31949; AAA64229.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63456 MW;  AE556302A9EEB99F CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 7
Q8UXR3
ID   Q8UXR3      PRELIMINARY;      PRT;   566 AA.
AC   Q8UXR3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Heamagglutinin.
GN   HA.
OS   Influenza A virus (A/swine/Potsdam/35/82(H3N2)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=183769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/swine/Potsdam/35/82;
RA   Groetzinger I., Suess J., Groetzinger C.;
RT   "Evolution of european human and porcine influenza viruses.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ252132; CAC81018.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63529 MW;  6AA44C84B4DDE68A CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 8
Q91MA7
ID   Q91MA7      PRELIMINARY;      PRT;   566 AA.
AC   Q91MA7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hemagglutinin.
OS   Influenza A virus (A/Hong Kong/1/68(H3N2)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=108859;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Hong Kong/1/68(H3N2);
RX   MEDLINE=21287244; PubMed=11371620;
```

```
RA   Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;
RT   "Pattern of mutation in the genome of influenza A virus on adaptation
RT   to increased virulence in the mouse lung: Identification of functional
RT   themes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AF348176; AAK51718.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63387 MW;  01BB0D465BE158E1 CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 9
Q9DHG0
ID   Q9DHG0      PRELIMINARY;      PRT;   566 AA.
AC   Q9DHG0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Haemagglutinin precursor.
OS   Influenza A virus H3N2.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=41857;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=clone 7a;
RA   Mohsin M.A., Morris S.J., Smith H., Sweet C.;
RT   "Influenza virus-induced apoptosis: a dual role for viral
RT   neuraminidase.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ289703; CAC18525.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT   SIGNAL        1     16       POTENTIAL.
SQ   SEQUENCE   566 AA;  63356 MW;  0BA681929300F72F CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 10
Q910M5
ID   Q910M5      PRELIMINARY;      PRT;   566 AA.
AC   Q910M5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hemagglutinin.
OS   Influenza A virus (A/Hong Kong/1/68(H3N2)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=108859;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Hong Kong/1/68(H3N2);
RX   MEDLINE=21287244; PubMed=11371620;
RA   Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;
RT   "Pattern of mutation in the genome of influenza A virus on adaptation
RT   to increased virulence in the mouse lung: Identification of functional
RT   themes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AF348179; AAK51721.1; -.
DR   EMBL; AF348177; AAK51719.1; -.
DR   EMBL; AF348178; AAK51720.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63530 MW;  7CB9F5BAF1E6E9F4 CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 11
Q67126
ID   Q67126      PRELIMINARY;      PRT;   566 AA.
AC   Q67126;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin.
GN   HA.
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Seal/MA/3984/92;
RX   MEDLINE=95146951; PubMed=7844533;
RA   Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT   "The appearance of H3 influenza viruses in seals.";
RL   J. Gen. Virol. 76:199-203(1995).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; L32024; AAA64228.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
```

```
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63441 MW;  590576CB4CEE7D08 CRC64;

  Query Match             93.7%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-10;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 12
Q9DXE3
ID   Q9DXE3      PRELIMINARY;      PRT;   301 AA.
AC   Q9DXE3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin (Fragment).
GN   HA1.
OS   Influenza A virus (A/Shorebird/Taiwan/31-4/99).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=140665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Shorebird/Taiwan/31-4/99;
RA   Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.;
RT   "Identification and subtyping of avian influenza virus by reverse
RT   transcription-polymerase chain reaction.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AF311750; AAG33016.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER       1      1
FT   NON_TER     301    301
SQ   SEQUENCE   301 AA;  32701 MW;  62A403758B764D57 CRC64;

  Query Match             93.0%;  Score 133;  DB 12;  Length 301;
  Best Local Similarity   95.7%;  Pred. No. 6.2e-10;
  Matches   22;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              ||||||||||||||||||:||||
Db        250 GLFGAIAGFIENGWEGMVDGWYG 272


RESULT 13
Q82499
ID   Q82499      PRELIMINARY;      PRT;   550 AA.
AC   Q82499;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinins HA1 and HA2 (Fragment).
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82/BS;
RA   Hartley C.A., Ward A.C., Anders E.M.;
RT   "Virulence of influenza virus for mice is associated with loss of
RT   oligosaccharide from the hemagglutinin molecule.";
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; U08859; AAA18782.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER       1      1
FT   CHAIN         1    328       HA1.
FT   CHAIN       330    550       HA2.
SQ   SEQUENCE   550 AA;  61772 MW;  50BD62B6BFE11FD8 CRC64;

  Query Match             92.3%;  Score 132;  DB 12;  Length 550;
  Best Local Similarity   95.7%;  Pred. No. 1.6e-09;
  Matches   22;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |:|||||||||||||||||||||
Db        330 GIFGAIAGFIENGWEGMIDGWYG 352


RESULT 14
Q82498
ID   Q82498      PRELIMINARY;      PRT;   550 AA.
AC   Q82498;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinins HA1 and HA2 (Fragment).
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82;
RA   Hartley C.A., Ward A.C., Anders E.M.;
RT   "Virulence of influenza virus for mice is associated with loss of
RT   oligosaccharide from the hemagglutinin molecule.";
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82;
RX   MEDLINE=88185444; PubMed=3356226;
RA   Nakajima S., Takeuchi Y., Nakajima K.;
RT   "Location on the evolutionary tree of influenza H3 haemagglutinin
RT   genes of Japanese strains isolated during the 1985-6 season.";
RL   Epidemiol. Infect. 100:301-310(1988).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; U08858; AAA18781.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER       1      1
FT   CHAIN         1    328       HA1.
FT   CHAIN       330    550       HA2.
SQ   SEQUENCE   550 AA;  61802 MW;  114413B1CE5A1F6A CRC64;

  Query Match             92.3%;  Score 132;  DB 12;  Length 550;
```

```
  Best Local Similarity   95.7%;  Pred. No. 1.6e-09;
  Matches   22;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |:|||||||||||||||||||||
Db        330 GIFGAIAGFIENGWEGMIDGWYG 352


RESULT 15
Q82753
ID   Q82753      PRELIMINARY;      PRT;   550 AA.
AC   Q82753;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Haemagglutinin (Fragment).
OS   Influenza virus.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   unclassified Orthomyxoviridae.
OX   NCBI_TaxID=11309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82/BS/ML10;
RX   MEDLINE=97300854; PubMed=9155874;
RA   Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;
RT   "Changes in the hemagglutinin molecule of influenza type A (H3N2)
RT   virus associated with increased virulence for mice.";
RL   Arch. Virol. 142:75-88(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82/BS/ML10;
RX   MEDLINE=97456249; PubMed=9311563;
RA   Ward A.C.;
RT   "Virulence of influenza A virus for mouse lung.";
RL   Virus Genes 14:187-194(1997).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; U08905; AAC79579.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER       1      1
FT   CHAIN         1    328       HAEMAGGLUTININ HA1.
FT   CHAIN       330    550       HAEMAGGLUTININ HA2.
SQ   SEQUENCE   550 AA;  61745 MW;  692A49DE678AC4BC CRC64;

  Query Match             92.3%;  Score 132;  DB 12;  Length 550;
  Best Local Similarity   95.7%;  Pred. No. 1.6e-09;
  Matches   22;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |:|||||||||||||||||||||
Db        330 GIFGAIAGFIENGWEGMIDGWYG 352

Search completed: January 30, 2004, 00:24:41
Job time : 26.7887 secs
```

```
OM protein - protein search, using sw model

Run on:          January 30, 2004, 00:13:12 ; Search time 19.2254 Seconds
                                              (without alignments)
                                              225.098 Million cell updates/sec

Title:           US-09-461-684C-4
Perfect score:   209
Sequence:        1 CEAAAAAEAAAAAEAAAAAE..........KKKKKKKKKKKKKKKKKKK 45

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_76:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	103	49.3	248	1	HSUR1P	histone H1, gonada
2	102	48.8	517	2	T49173	hypothetical prote
3	101.5	48.6	442	2	T39683	zuotin-like protei
4	100	47.8	215	2	I52523	nucleoporin p62 ho
5	100	47.8	380	2	T46395	hypothetical prote
6	98	46.9	206	2	S09388	histone H1 - sea u
7	96	45.9	433	2	S25194	zuotin - yeast (Sa
8	95	45.5	166	2	T18513	hypothetical prote
9	91	43.5	229	2	JC7219	nuclear protein SR
10	90.5	43.3	383	2	A86315	F2H15.19 protein -
11	90	43.1	153	2	S59591	histone H2B (clone
12	90	43.1	483	2	F71619	hypothetical prote
13	90	43.1	529	2	T50609	hypothetical prote
14	89	42.6	208	2	T23778	histone H1.1 - Cae
15	89	42.6	265	2	S19113	cgcr-4 protein - C
16	88	42.1	392	2	T15755	hypothetical prote
17	88	42.1	409	2	T18726	hypothetical prote
18	87	41.6	107	2	C86477	protein F15O4.29 [
19	87	41.6	565	2	T47775	hypothetical prote
20	86	41.1	347	2	E83525	TolA protein PA097
21	86	41.1	455	2	A87913	protein B0205.10 [
22	85.5	40.9	1280	2	G96796	hypothetical prote
23	85	40.7	102	2	D64363	ribosomal protein
24	85	40.7	153	2	S59587	histone H2B (clone
25	85	40.7	241	2	JN0748	histone H1-II - Vo
26	85	40.7	560	2	T06377	SAR DNA-binding pr
27	85	40.7	849	2	T01286	probable RNA-bindi
28	84.5	40.4	228	2	T36379	probable peptidase
29	84	40.2	409	2	T24543	hypothetical prote
30	84	40.2	425	2	T18723	hypothetical prote
31	84	40.2	441	2	A48455	acidic phosphoprot
32	84	40.2	1560	2	T42727	proliferation pote
33	84	40.2	1701	2	T09127	probable erythrocy
34	83	39.7	107	2	A75031	lsu ribosomal prot
35	83	39.7	146	2	AI3340	hypothetical prote
36	83	39.7	320	2	S61586	probable membrane
37	83	39.7	421	2	JV0057	tolA protein - Esc
38	83	39.7	625	2	T39019	probable mitochond
39	82	39.2	153	2	S59125	histone H2B [valid
40	82	39.2	474	2	T38485	centromere/microtu
41	81.5	39.0	628	2	T08942	proton pump intera
42	81	38.8	111	2	F71216	probable ribosomal
43	81	38.8	218	2	JC7220	nuclear protein SR
44	80.5	38.5	1002	2	S70292	FUN12 protein - ye
45	80	38.3	111	2	E72524	probable ribosomal

ALIGNMENTS

```
RESULT 1
HSUR1P
histone H1, gonadal - sea urchin (Parechinus angulosus)
C;Species: Parechinus angulosus (angulate urchin)
C;Date: 31-Mar-1980 #sequence_revision 31-Mar-1980 #text_change 16-Feb-1997
C;Accession: A91090; A91091; A02586
R;Strickland, W.N.; Strickland, M.; de Groot, P.C.; von Holt, C.; Wittmann-Liebold, B.
Eur. J. Biochem. 104, 559-566, 1980
A;Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus angu
rminal cyanogen bromide peptides.
A;Reference number: A91090; MUID:80156831; PMID:6767609
A;Contents: sequence of residues 1-84
A;Accession: A91090
A;Molecule type: protein
A;Residues: 1-248 <STR>
R;Strickland, W.N.; Strickland, M.; Brandt, W.F.; von Holt, C.; Lehmann, A.; Wittmann-Lie
Eur. J. Biochem. 104, 567-578, 1980
A;Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus angu
A;Reference number: A91091; MUID:80156832; PMID:7363905
A;Accession: A91091
A;Molecule type: protein
A;Residues: 80-248 <ST2>
A;Note: 144-Arg was also found
C;Superfamily: histone H1
C;Keywords: DNA binding; nucleosome; sperm

  Query Match            49.3%;  Score 103;  DB 1;  Length 248;
  Best Local Similarity  59.5%;  Pred. No. 0.086;
  Matches   25;  Conservative    4;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 43
                :|||  :|| | : ||||:  |||| || || |||  || ||
Db        146 KAAAKRKAALAKKKAAAAKRKAAAKAKKAKKPKKKAAKKAKK 187


RESULT 2
T49173
hypothetical protein T20N10.250 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
C;Accession: T49173
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25017
A;Accession: T49173
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-517 <DAN>
A;Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
A;Experimental source: cultivar Columbia; BAC clone T20N10
C;Genetics:
A;Gene: ATSP:T20N10.250
```

```
A;Map position: 3
A;Introns: 312/3; 359/3; 444/3
C;Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

  Query Match             48.8%;  Score 102;  DB 2;  Length 517;
  Best Local Similarity   80.8%;  Pred. No. 0.17;
  Matches   21;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         20 EAAAAAKKKKKKKKKKKKKKKKKKKK 45
              |     ||||||||||||||||||||
Db        440 ERVGKKKKKKKKKKKKKKKKKKKKKK 465

RESULT 3
T39683
zuotin-like protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T39683; T40195
R;Oliver, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, March 1998
A;Reference number: Z21869
A;Accession: T39683
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-124 <OLI>
A;Cross-references: EMBL:AL049489; PIDN:CAB39796.1; GSPDB:GN00067; SPDB:SPBC1778.01c
A;Experimental source: strain 972h-; cosmid c1778
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21910
A;Accession: T40195
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 89-442 <WOO>
A;Cross-references: EMBL:Z97992; PIDN:CAB10796.1; GSPDB:GN00067; SPDB:SPBC30D10.01
A;Experimental source: strain 972h-; cosmid c30D10
C;Genetics:
A;Gene: SPDB:SPBC1778.01c; SPDB:SPBC30D10.01
A;Map position: 2

  Query Match             48.6%;  Score 101.5;  DB 2;  Length 442;
  Best Local Similarity   56.0%;  Pred. No. 0.17;
  Matches   28;  Conservative    3;  Mismatches   12;  Indels    7;  Gaps    1;

Qy          2 EAAAAA-------EAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKK 44
              ||||||       |  || |||| | |||| || |: ||| :|: ||  |
Db        309 EAAAAAQKKKEEEERRAAEEAAAKASAAAANKKAKEDKKKAQKRDKKVVK 358

RESULT 4
I52523
nucleoporin p62 homolog - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
C;Accession: I52523
R;Wang, Z.Q.; Akmal, K.M.; Kim, K.H.
Biol. Reprod. 51, 1022-1030, 1994
A;Title: An unusual nucleoporin-related messenger ribonucleic acid is present in the ger
A;Reference number: I52523; MUID:95151924; PMID:7849178
A;Accession: I52523
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-215 <RES>
A;Cross-references: GB:S75997; NID:g913245; PIDN:AAB33384.1; PID:g913246
A;Experimental source: testis

  Query Match             47.8%;  Score 100;  DB 2;  Length 215;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 KKKKKKKKKKKKKKKKKKKK 45
              ||||||||||||||||||||
Db         35 KKKKKKKKKKKKKKKKKKKK 54

RESULT 5
T46395
hypothetical protein DKFZp434I1120.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46395
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46395
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-380 <AAA>
A;Cross-references: EMBL:AL137556
A;Experimental source: adult testis; clone DKFZp434I1120
C;Genetics:
A;Note: DKFZp434I1120.1

  Query Match             47.8%;  Score 100;  DB 2;  Length 380;
  Best Local Similarity   100.0%;  Pred. No. 0.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 KKKKKKKKKKKKKKKKKKKK 45
              ||||||||||||||||||||
Db        355 KKKKKKKKKKKKKKKKKKKK 374

RESULT 6
S09388
histone H1 - sea urchin (Parechinus angulosus)
C;Species: Parechinus angulosus (angulate urchin)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C;Accession: S09388
R;Hill, C.S.; Martin, S.R.; Thomas, J.O.
EMBO J. 8, 2591-2599, 1989
A;Title: A stable alpha-helical element in the carboxy-terminal domain of free and chroma
A;Reference number: S09388; MUID:90060019; PMID:2583125
A;Accession: S09388
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-206 <HIL>
C;Superfamily: histone H1
C;Keywords: chromosomal protein

  Query Match             46.9%;  Score 98;  DB 2;  Length 206;
  Best Local Similarity   55.8%;  Pred. No. 0.18;
  Matches   24;  Conservative    4;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKK 44
               :|||  :|| | : ||||: || | || || |||   || ||
Db        140 KAAAKRKAALAKKKAAAAKRKAATKAKKAKKPKKKTAAKKAKK 182

RESULT 7
S25194
zuotin - yeast (Saccharomyces cerevisiae)
N;Alternate names: probable Z-DNA-binding protein; protein G9554; protein YGR285c
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 21-Jul-2000
C;Accession: S25194; S64620; S19066
R;Zhang, S.; Lockshin, C.; Herbert, A.; Winter, E.; Rich, A.
EMBO J. 11, 3787-3796, 1992
A;Title: Zuotin, a putative Z-DNA binding protein in Saccharomyces cerevisiae.
A;Reference number: S25194; MUID:93010971; PMID:1396572
A;Accession: S25194
A;Molecule type: DNA
A;Residues: 1-433 <ZHA>
A;Cross-references: EMBL:X63612; NID:g4836; PIDN:CAA45156.1; PID:g4837
A;Note: part of this sequence, including the amino end of the mature protein, was confirm
R;Voet, M.; Volckaert, G.
```

```
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64611
A;Accession: S64620
A;Molecule type: DNA
A;Residues: 1-433 <VOE>
A;Cross-references: EMBL:Z73070; NID:g1323520; PID:e243256; PID:g1323521; MIPS:YGR285c
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:ZUO1
A;Cross-references: SGD:S0003517; MIPS:YGR285c
A;Map position: 7R
C;Superfamily: dnaJ amino-terminal homology
C;Keywords: DNA binding; nucleus
F;97-168/Domain: dnaJ amino-terminal homology <DNJ>

  Query Match             45.9%;  Score 96;  DB 2;  Length 433;
  Best Local Similarity   52.3%;  Pred. No. 0.44;
  Matches   23;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKK 45
                || |  ||| | |:|  : |:| |:||  ||| |:  |  ||| |:
Db        313 EAKAKAEAEAKAKAESEAKANASAKADKKKAKEAAKAAKKKNKR 356

RESULT 8
T18513
hypothetical protein C0845c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T18513
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: Z18935
A;Accession: T18513
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-166 <LAW>
A;Cross-references: EMBL:Z98551; PIDN:CAB11123.2
C;Genetics:
A;Map position: 3
A;Introns: 19/1
A;Note: C0845c

  Query Match             45.5%;  Score 95;  DB 2;  Length 166;
  Best Local Similarity   82.6%;  Pred. No. 0.27;
  Matches   19;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         23 AAAKKKKKKKKKKKKKKKKKKKK 45
                |  ||||||:||||:|||||||
Db         34 ALIKKKKKKEKKKKRKKKKKKKK 56

RESULT 9
JC7219
nuclear protein SR-25 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C;Accession: JC7219
R;Sasahara, K.; Yamaoka, T.; Moritani, M.; Tanaka, M.; Iwahana, H.; Yoshimoto, K.; Miyag
Biochem. Biophys. Res. Commun. 269, 444-450, 2000
A;Title: Molecular cloning and expression analysis of a putative nuclear protein, SR-25.
A;Reference number: JC7219; MUID:20175222; PMID:10708573
A;Accession: JC7219
A;Molecule type: mRNA
A;Residues: 1-229 <SAS>
A;Cross-references: DDBJ:AB035383; NID:g7619895; PIDN:BAA94743.1; PID:g7619896
A;Experimental source: MIN6 cell line
C;Comment: This protein is a highly hydrophilic nuclear protein with a serine-arginine r
A splicing factors.
C;Keywords: nucleus; RNA processing

  Query Match             43.5%;  Score 91;  DB 2;  Length 229;
  Best Local Similarity   34.9%;  Pred. No. 0.69;
  Matches   15;  Conservative   19;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKK 45
              ::: : ::::: :|::: ::   ||: | |:||:||||||:||
Db         68 SSSTSSSSSSSSASSSSSSDGRKKRAKHKEKKRKKKKKKRKK 110

RESULT 10
A86315
F2H15.19 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C;Accession: A86315
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86315
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-383 <STO>
A;Cross-references: GB:AE005172; NID:g9665074; PIDN:AAF97276.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

  Query Match             43.3%;  Score 90.5;  DB 2;  Length 383;
  Best Local Similarity   48.9%;  Pred. No. 1.1;
  Matches   23;  Conservative    5;  Mismatches   16;  Indels    3;  Gaps    1;

Qy          1 CEAAAAA---EAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 44
              | |||||    | :  :     :|:  |   :| ||||| | |||||| |
Db         80 CYAAAAAAAQEVSVPSVIRYSADLDARIRKDKKKKKHKHKKKKKKNK 126

RESULT 11
S59591
histone H2B (clone CH-IV) - Chlamydomonas reinhardtii
C;Species: Chlamydomonas reinhardtii
C;Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 23-Jul-1999
C;Accession: S59591
R;Fabry, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius, T.; Schmitt, R.
Curr. Genet. 28, 333-345, 1995
A;Title: The organization structure and regulatory elements of Chlamydomonas histone gene
A;Reference number: S59581; MUID:96120862; PMID:8590479
A;Accession: S59591
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-153 <FAB>
A;Cross-references: EMBL:U16726; NID:g571479; PIDN:AAA98454.1; PID:g571482
A;Note: the authors did not translate the codon for residue 1
C;Superfamily: histone H2B

  Query Match             43.1%;  Score 90;  DB 2;  Length 153;
  Best Local Similarity   51.0%;  Pred. No. 0.62;
  Matches   25;  Conservative    2;  Mismatches   16;  Indels    6;  Gaps    1;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKK------KKKKKKKK 45
              | | ||| | |||   || ||   ||:  ||  |        :|| |||||
Db         12 AEAGAEAPAKAEAKPKAEKAAKKAKKEPSKKAAKEPKGDGEKKDKKKKK 60

RESULT 12
F71619
hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
```

C;Accession: F71619
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: F71619
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-483 <GAR>
A;Cross-references: GB:AE001382; GB:AE001362; NID:g3845130; PIDN:AAC71836.1; PID:g384513
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0235w

```
  Query Match             43.1%;  Score 90;  DB 2;  Length 483;
  Best Local Similarity   85.0%;  Pred. No. 1.4;
  Matches   17;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 KKKKKKKKKKKKKKKKKKKK 45
              ||||||||:|||::||||||
Db        449 KKKKKKKKRKKKERKKKKKK 468
```

RESULT 13
T50609
hypothetical protein DKFZp761B2423.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T50609
R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, June 2000
A;Reference number: Z25143
A;Accession: T50609
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-529 <AAA>
A;Cross-references: EMBL:AL359564
A;Experimental source: adult amygdala; clone DKFZp761B2423
C;Genetics:
A;Note: DKFZp761B2423.1

```
  Query Match             43.1%;  Score 90;  DB 2;  Length 529;
  Best Local Similarity   45.0%;  Pred. No. 1.5;
  Matches   18;  Conservative   11;  Mismatches   11;  Indels    0;  Gaps    0;

Qy          5 AAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKK 44
              : ::  : : :|::||:  : ||||||:||| || || ||
Db        444 SGSQTRSHSSSASSAESQDSKKKKKKKEKKKHKKHKKHKK 483
```

RESULT 14
T23778
histone H1.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T23778; T42231; S09130; S01817
R;Percy, C.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19798
A;Accession: T23778
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-208 <WIL>
A;Cross-references: EMBL:Z79603; PIDN:CAB01892.1; GSPDB:GN00028; CESP:M163.3
A;Experimental source: clone M163
R;Jedrusik, M.; Schulze, E.
submitted to the EMBL Data Library, August 1997
A;Description: The histone H1 complement of Caenorhabditis elegans.
A;Reference number: Z22091
A;Accession: T42231
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-208 <JED>
A;Cross-references: EMBL:AF017810; PIDN:AAB70665.1
R;Sanicola, M.; Ward, S.; Childs, G.; Emmons, S.W.
J. Mol. Biol. 212, 259-268, 1990
A;Title: Identification of a Caenorhabditis elegans histone H1 gene family. Characterizat
A;Reference number: S09130; MUID:90204554; PMID:1969492
A;Accession: S09130
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-43,'T',45-83,'H',85-100,'R',102-208 <SAN>
A;Cross-references: GB:X53277; NID:g10885; PIDN:CAA37372.1; PID:g10886
R;Vanfleteren, J.R.; van Bun, S.M.; van Beeumen, J.J.
Biochem. J. 255, 647-652, 1988
A;Title: The primary structure of the major isoform (H1.1) of histone H1 from the nematoc
A;Reference number: S01817; MUID:89076229; PMID:3202838
A;Accession: S01817
A;Molecule type: protein
A;Residues: 2-43,'T',45-100,'K',102-208 <VAN>
C;Genetics:
A;Gene: CESP:M163.3; his-24
A;Map position: X
A;Introns: 79/3
C;Superfamily: histone H1
C;Keywords: blocked amino end; chromosomal protein; DNA binding; nucleosome; nucleus
F;2-208/Product: histone H1.1 #status predicted <MAT>
F;2/Modified site: blocked amino end (Ser) (in mature form) (probably acetylated) #statu

```
  Query Match             42.6%;  Score 89;  DB 2;  Length 208;
  Best Local Similarity   53.5%;  Pred. No. 0.92;
  Matches   23;  Conservative    4;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKK 44
              | ||||:  |||:  |||:  |||||   :|| ||    | ||
Db        113 EKAAAAKKPAAAKKPAAAKKPAAAKKATGEKKAKKPAAAKPKK 155
```

RESULT 15
S19113
cgcr-4 protein - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C;Accession: S19113; S14466
R;Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.
Plant Mol. Biol. 18, 143-146, 1992
A;Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements
A;Reference number: S19113; MUID:92119224; PMID:1731966
A;Accession: S19113
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-265 <WAK>
A;Cross-references: EMBL:X17208; NID:g18136; PIDN:CAA35080.1; PID:g18137
C;Genetics:
A;Gene: cgcr-4

```
  Query Match             42.6%;  Score 89;  DB 2;  Length 265;
  Best Local Similarity   68.8%;  Pred. No. 1.1;
  Matches   22;  Conservative    3;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKK 34
              ||| | |||||||||||:| |||: | |:  |
Db        176 AAAEARAAAAAEAAAAAKARAAAEAKAKEADK 207
```

Search completed: January 30, 2004, 00:26:22
Job time : 20.2254 secs

```
OM protein - protein search, using sw model

Run on:          January 29, 2004, 23:58:27 ; Search time 10.1408 Seconds
                                              (without alignments)
                                              208.681 Million cell updates/sec

Title:           US-09-461-684C-4
Perfect score:   209
Sequence:        1 CEAAAAAEAAAAAEAAAAAE..........KKKKKKKKKKKKKKKKKK 45

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:     127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	103	49.3	248	1	H1_PARAN	P02256 parechinus
2	101.5	48.6	442	1	YOI1_SCHPO	Q9y718 schizosacch
3	96	45.9	433	1	ZUO1_YEAST	P32527 saccharomyc
4	93	44.5	332	1	FA9A_HUMAN	Q8izu1 homo sapien
5	90	43.1	153	1	H2B4_CHLRE	P54347 chlamydomon
6	89	42.6	207	1	H11_CAEEL	P10771 caenorhabdi
7	86	41.1	347	1	TOLA_PSEAE	P50600 pseudomonas
8	85	40.7	102	1	RL12_METJA	P54048 methanococc
9	85	40.7	153	1	H2B3_CHLRE	P54346 chlamydomon
10	85	40.7	240	1	H12_VOLCA	Q08865 volvox cart
11	84	40.2	441	1	PHPA_PLACH	Q02752 plasmodium
12	84	40.2	726	1	BRD3_HUMAN	Q15059 homo sapien
13	83	39.7	107	1	RL12_PYRAB	Q9uxs6 pyrococcus
14	83	39.7	320	1	YD33_YEAST	Q12117 saccharomyc
15	83	39.7	421	1	TOLA_ECOLI	P19934 escherichia
16	82	39.2	153	1	H2B1_CHLRE	P50565 chlamydomon
17	82	39.2	474	1	CBF5_SCHPO	O14007 schizosacch
18	81	38.8	108	1	RL12_PYRHO	O57705 pyrococcus
19	81	38.8	487	1	CBF5_ASPFU	O43102 aspergillus
20	80.5	38.5	534	1	NOP5_RAT	Q9qz86 rattus norv
21	80.5	38.5	1002	1	IF2P_YEAST	P39730 saccharomyc
22	80	38.3	111	1	RL12_AERPE	Q9y9w9 aeropyrum p
23	80	38.3	237	1	H1E_CHITE	P40278 chironomus
24	80	38.3	508	1	NO60_DROME	O44081 drosophila
25	79.5	38.0	156	1	H2B2_CHLRE	P54345 chlamydomon
26	79.5	38.0	511	1	NOP5_YEAST	Q12499 saccharomyc
27	79.5	38.0	2231	1	SEN1_YEAST	Q00416 saccharomyc
28	79	37.8	109	1	RLA1_TRYCR	P26643 trypanosoma
29	79	37.8	171	1	H1_ECHCR	P02257 echinolampa
30	79	37.8	260	1	H11_VOLCA	Q08864 volvox cart
31	78.5	37.6	1391	1	MST2_DROHY	Q08696 drosophila
32	78	37.3	134	1	H1_EUPEU	P32103 euplotes eu
33	78	37.3	210	1	H1_LYTPI	P06144 lytechinus
34	78	37.3	218	1	H11R_CHICK	P08288 gallus gall
35	78	37.3	218	1	H12_RAT	P15865 rattus norv
36	78	37.3	218	1	H14_HUMAN	P10412 homo sapien
37	78	37.3	218	1	H14_MOUSE	P43274 mus musculu
38	78	37.3	372	1	TOLA_HAEIN	P44678 haemophilus
39	78	37.3	414	1	Y694_METJA	Q58105 methanococc
40	78	37.3	1220	1	IF2P_HUMAN	O60841 homo sapien
41	77.5	37.1	671	1	CHS5_YEAST	Q12114 saccharomyc
42	77	36.8	101	1	RL12_METTL	O52706 methanococc
43	77	36.8	113	1	RLA1_DICDI	P22684 dictyosteli
44	77	36.8	211	1	H12_MOUSE	P15864 mus musculu
45	77	36.8	220	1	H13_MOUSE	P43277 mus musculu

ALIGNMENTS

```
RESULT 1
H1_PARAN
ID   H1_PARAN       STANDARD;      PRT;   248 AA.
AC   P02256;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1, gonadal.
OS   Parechinus angulosus (Angulate sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;
OC   Parechinus.
OX   NCBI_TaxID=7658;
RN   [1]
RP   SEQUENCE OF 1-84.
RX   MEDLINE=80156831; PubMed=6767609;
RA   Strickland W.N., Strickland M., de Groot P.C., von Holt C.,
RA   Wittmann-Liebold B.;
RT   "The primary structure of histone H1 from sperm of the sea urchin
RT   Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT   protein and the sequence of amino acids in the four N-terminal
RT   cyanogen bromide peptides.";
RL   Eur. J. Biochem. 104:559-566(1980).
RN   [2]
RP   SEQUENCE OF 80-248.
RX   MEDLINE=80156832; PubMed=7363905;
RA   Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,
RA   Wittmann-Liebold B.;
RT   "The primary structure of histone H1 from sperm of the sea urchin
RT   Parechinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT   the entire primary structure.";
RL   Eur. J. Biochem. 104:567-578(1980).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: SPERM.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR   PIR; A91090; HSUR1P.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Sperm.
FT   VARIANT     144    144       K -> R.
SQ   SEQUENCE   248 AA;  26387 MW;  1B25B3F136541947 CRC64;

  Query Match             49.3%;  Score 103;  DB 1;  Length 248;
  Best Local Similarity   59.5%;  Pred. No. 0.019;
  Matches   25;  Conservative    4;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 43
```

```
             :|||  :|| | : ||||:  |||| || || |||  || ||
Db       146 KAAAKRKAALAKKKAAAAKRKAAAKAKKAKKPKKKAAKKAKK 187

RESULT 2
YOI1_SCHPO
ID   YOI1_SCHPO     STANDARD;      PRT;   442 AA.
AC   Q9Y7I8; O14347;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hypothetical J-domain protein C1778.01c in chromosome II.
GN   SPBC1778.01C OR SPBC30D10.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SIMILARITY: Contains 1 J domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL049489; CAB39796.1; -.
DR   EMBL; Z97992; CAB10796.1; -.
DR   PIR; T39683; T39683.
DR   HSSP; P25685; 1HDJ.
DR   GeneDB_SPombe; SPBC1778.01c; -.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00226; DnaJ; 1.
DR   SMART; SM00271; DnaJ; 1.
DR   PROSITE; PS00636; DNAJ_1; 1.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   Hypothetical protein; Chaperone.
FT   DOMAIN       97    167       J-DOMAIN.
FT   DOMAIN      294    342       ALA/LYS-RICH.
SQ   SEQUENCE   442 AA;  50209 MW;  F4EC924871B7318B CRC64;

  Query Match             48.6%;  Score 101.5;  DB 1;  Length 442;
  Best Local Similarity   56.0%;  Pred. No. 0.042;
  Matches   28;  Conservative    3;  Mismatches   12;  Indels    7;  Gaps    1;

Qy         2 EAAAAA-------EAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKK 44
             ||||||       |  || |||| | |||| || |: ||| :|: ||  |
Db       309 EAAAAAQKKKEEEERRAAEEAAAKASAAAANKKAKEDKKKAQKRDKKVVK 358

RESULT 3
ZUO1_YEAST
ID   ZUO1_YEAST     STANDARD;      PRT;   433 AA.
AC   P32527;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zuotin.
GN   ZUO1 OR YGR285C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC   STRAIN=20B-12;
RX   MEDLINE=93010971; PubMed=1396572;
RA   Zhang S., Lockshin C., Herbert A., Winter E., Rich A.;
RT   "Zuotin, a putative Z-DNA binding protein in Saccharomyces
RT   cerevisiae.";
RL   EMBO J. 11:3787-3796(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=97245295; PubMed=9090054;
RA   Volckaert G., Voet M., Robben J.;
RT   "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the
RT   right arm of chromosome VII from Saccharomyces cerevisiae carrying
RT   the MAL1 locus reveals 15 complete open reading frames, including
RT   ZUO1, BGL2 and BIO2 genes and an ABC transporter gene.";
RL   Yeast 13:251-259(1997).
CC   -!- FUNCTION: Z-DNA BINDING PROTEIN. COULD BE INVOLVED IN
CC       CHROMOSOME ORGANIZATION.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 1 J domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X63612; CAA45156.1; -.
DR   EMBL; Z73070; CAA97317.1; -.
DR   PIR; S25194; S25194.
DR   HSSP; P08622; 1BQZ.
DR   SGD; S0003517; ZUO1.
DR   GO; GO:0005829; C:cytosol; NAS.
DR   GO; GO:0005840; C:ribosome; NAS.
DR   GO; GO:0003754; F:chaperone activity; NAS.
DR   GO; GO:0006457; P:protein folding; NAS.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00226; DnaJ; 1.
DR   SMART; SM00271; DnaJ; 1.
DR   PROSITE; PS00636; DNAJ_1; 1.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   Chaperone; DNA-binding; Nuclear protein.
FT   DOMAIN       98    170       J-DOMAIN.
FT   DOMAIN      306    357       ALA/LYS-RICH.
SQ   SEQUENCE   433 AA;  49019 MW;  0AA76BC11D3C7DAB CRC64;

  Query Match             45.9%;  Score 96;  DB 1;  Length 433;
```

```
  Best Local Similarity   52.3%;  Pred. No. 0.11;
  Matches   23;  Conservative    6;  Mismatches   15;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKK 45
              || | ||| | |:| : |:| |:|| ||| |: | ||| |:
Db        313 EAKAKAEAEAKAKAESEAKANASAKADKKKAKEAAKAAKKKNKR 356


RESULT 4
FA9A_HUMAN
ID   FA9A_HUMAN     STANDARD;      PRT;   332 AA.
AC   Q8IZU1;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Protein FAM9A.
GN   FAM9A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
RX   MEDLINE=22202142; PubMed=12213195;
RA   Martinez-Garay I., Jablonka S., Sutajova M., Steuernagel P., Gal A.,
RA   Kutsche K.;
RT   "A new gene family (FAM9) of low-copy repeats in Xp22.3 expressed
RT   exclusively in testis: implications for recombinations in this
RT   region.";
RL   Genomics 80:259-267(2002).
CC   -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC   -!- TISSUE SPECIFICITY: Expressed exclusively in testis.
CC   -!- SIMILARITY: Belongs to the FAM9 family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF494343; AAN07162.1; -.
DR   Genew; HGNC:18403; FAM9A.
KW   Nuclear protein.
FT   DOMAIN      180    275       GLU-RICH.
FT   DOMAIN      194    214       POLY-ALA.
FT   DOMAIN      252    258       POLY-GLY.
SQ   SEQUENCE   332 AA;  37339 MW;  92F22EC36038229C CRC64;

  Query Match             44.5%;  Score 93;  DB 1;  Length 332;
  Best Local Similarity   44.2%;  Pred. No. 0.16;
  Matches   19;  Conservative   17;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKK 45
              ||||||||||||||||       : :::::|:::::|::::::
Db        197 AAAAAEAAAAAEAAAAAAEVIVVEDEEEEKEEEEEKEEEEEE 239


RESULT 5
H2B4_CHLRE
ID   H2B4_CHLRE     STANDARD;      PRT;   153 AA.
AC   P54347;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Histone H2B-IV.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96120862; PubMed=8590479;
RA   Fabry S., Mueller K., Lindauer A., Park P.B., Cornelius T.,
RA   Schmitt R.;
RT   "The organization structure and regulatory elements of Chlamydomonas
RT   histone genes reveal features linking plant and animal genes.";
RL   Curr. Genet. 28:333-345(1995).
CC   -!- SUBUNIT: The nucleosome is an octamer containing two molecules
CC       each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
CC       BP of DNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the histone H2B family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U16726; AAA98454.1; -.
DR   PIR; S59591; S59591.
DR   InterPro; IPR004822; Histone_core.
DR   InterPro; IPR000558; Histone_H2B.
DR   Pfam; PF00125; histone; 1.
DR   PRINTS; PR00621; HISTONEH2B.
DR   ProDom; PD000497; Histone_H2B; 1.
DR   SMART; SM00427; H2B; 1.
DR   PROSITE; PS00357; HISTONE_H2B; 1.
KW   Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW   Multigene family.
SQ   SEQUENCE   153 AA;  16587 MW;  B09C1ED804E1F485 CRC64;

  Query Match             43.1%;  Score 90;  DB 1;  Length 153;
  Best Local Similarity   51.0%;  Pred. No. 0.14;
  Matches   25;  Conservative    2;  Mismatches   16;  Indels    6;  Gaps    1;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKK------KKKKKKKKK 45
              | | ||| | |||   || ||   ||:  || |       :|| |||||
Db         12 AEAGAEAPAKAEAKPKAEKAAKKAKKEPSKKAAKEPKGDGEKKDKKKKK 60


RESULT 6
H11_CAEEL
ID   H11_CAEEL      STANDARD;      PRT;   207 AA.
AC   P10771;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Histone H1.1.
GN   HIS-24.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90204554; PubMed=1969492;
RA   Sanicola M., Ward S., Childs G., Emmons S.W.;
RT   "Identification of a Caenorhabditis elegans histone H1 gene family.
RT   Characterization of a family member containing an intron and encoding
RT   a poly(A)+ mRNA.";
RL   J. Mol. Biol. 212:259-268(1990).
RN   [2]
RP   SEQUENCE.
RC   STRAIN=Bristol N2;
RX   MEDLINE=89076229; PubMed=3202838;
RA   Vanfleteren J.R., van Bun S.M., van Beeumen J.J.;
RT   "The primary structure of the major isoform (H1.1) of histone H1 from
RT   the nematode Caenorhabditis elegans.";
RL   Biochem. J. 255:647-652(1988).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
```

```
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- MISCELLANEOUS: THERE ARE 2 FORMS OF H1 IN THIS NEMATODE: H1.1 IS
CC       THE MAJOR FORM.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X53277; CAA37372.1; -.
DR   HSSP; P02259; 1HST.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Acetylation.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION (PROBABLE).
FT   DOMAIN       36    112       GLOBULAR.
FT   CONFLICT     83     83       H -> L (IN REF. 2).
FT   CONFLICT    100    100       R -> K (IN REF. 2).
SQ   SEQUENCE   207 AA;  21314 MW;  7802EA9AA4F36F6F CRC64;

  Query Match             42.6%;  Score 89;  DB 1;  Length 207;
  Best Local Similarity   53.5%;  Pred. No. 0.23;
  Matches   23;  Conservative    4;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 44
                | ||||: |||: |||: |||||    :|| ||    | ||
Db        112 EKAAAAKKPAAAKKPAAAKKPAAAKKATGEKKAKKPAAAKPKK 154


RESULT 7
TOLA_PSEAE
ID   TOLA_PSEAE     STANDARD;      PRT;   347 AA.
AC   P50600;
DT   01-OCT-1996 (Rel. 34, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   TolA protein.
GN   TOLA OR PA0971.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PAO;
RX   MEDLINE=97113525; PubMed=8955385;
RA   Dennis J.J., Lafontaine E.R., Sokol P.A.;
RT   "Identification and characterization of the tolQRA genes of
RT   Pseudomonas aeruginosa.";
RL   J. Bacteriol. 178:7059-7068(1996).
RN   [2]
RP   REVISIONS TO N-TERMINUS.
RA   Duan K., Sokol P.A.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC       (Potential).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U39558; AAC44660.2; -.
DR   EMBL; AE004530; AAG04360.1; -.
DR   PIR; E83525; E83525.
DR   InterPro; IPR006260; TonB_C.
DR   TIGRFAMs; TIGR01352; tonB_Cterm; 1.
KW   Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
KW   Complete proteome.
FT   DOMAIN        1     16       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     17     37       POTENTIAL.
FT   DOMAIN       38    347       PERIPLASMIC (POTENTIAL).
FT   DOMAIN      209    216       POLY-ALA.
SQ   SEQUENCE   347 AA;  37935 MW;  EEDD4B04AA095945 CRC64;

  Query Match             41.1%;  Score 86;  DB 1;  Length 347;
  Best Local Similarity   51.2%;  Pred. No. 0.61;
  Matches   22;  Conservative    5;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 44
              ||  ||||  | ||  |||| || :||:     ||: : : |||
Db        124 EAQKAAEAKKADEAKKAAEAKAAEQKKQADIAKKRAEDEAKKK 166


RESULT 8
RL12_METJA
ID   RL12_METJA     STANDARD;      PRT;   102 AA.
AC   P54048;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   50S ribosomal protein L12P.
GN   RPL12P OR MJ0508.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC       INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC       ACCURATE TRANSLATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC   --------------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U67500; AAB98498.1; -.
DR   PIR; D64363; D64363.
DR   TIGR; MJ0508; -.
DR   InterPro; IPR001813; 60s_ribosomal.
DR   Pfam; PF00428; 60s_ribosomal; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   102 AA;  10363 MW;  35306FFDE967C52C CRC64;

  Query Match             40.7%;  Score 85;  DB 1;  Length 102;
  Best Local Similarity   52.6%;  Pred. No. 0.26;
  Matches   20;  Conservative    8;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKK 39
              || | |     ||  |||| |||| :||:::||::||::
Db         52 EAIANAAMPVAAAPAAAAPAAAAEEKKEEEKKEEKKEE 89


RESULT 9
H2B3_CHLRE
ID   H2B3_CHLRE     STANDARD;      PRT;   153 AA.
AC   P54346;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Histone H2B-III.
OS   Chlamydomonas reinhardtii.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Chlamydomonas.
OX   NCBI_TaxID=3055;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96120862; PubMed=8590479;
RA   Fabry S., Mueller K., Lindauer A., Park P.B., Cornelius T.,
RA   Schmitt R.;
RT   "The organization structure and regulatory elements of Chlamydomonas
RT   histone genes reveal features linking plant and animal genes.";
RL   Curr. Genet. 28:333-345(1995).
CC   -!- SUBUNIT: The nucleosome is an octamer containing two molecules
CC       each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
CC       BP of DNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the histone H2B family.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U16725; AAA98450.1; -.
DR   PIR; S59587; S59587.
DR   InterPro; IPR004822; Histone_core.
DR   InterPro; IPR000558; Histone_H2B.
DR   Pfam; PF00125; histone; 1.
DR   PRINTS; PR00621; HISTONEH2B.
DR   ProDom; PD000497; Histone_H2B; 1.
DR   SMART; SM00427; H2B; 1.
DR   PROSITE; PS00357; HISTONE_H2B; 1.
KW   Chromosomal protein; Nucleosome core; Nuclear protein; DNA-binding;
KW   Multigene family.
SQ   SEQUENCE   153 AA;  16557 MW;  2092413E04E1F49C CRC64;

  Query Match             40.7%;  Score 85;  DB 1;  Length 153;
  Best Local Similarity   49.0%;  Pred. No. 0.37;
  Matches   24;  Conservative    2;  Mismatches   17;  Indels    6;  Gaps    1;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKK------KKKKKKKKK 45
              | | ||| | |||   || |     ||:  ||  |        :|| |||||
Db         12 AEAGAEAPAKAEAKPKAEKAGKKAKKEPAKKAAKEPKGDGEKKDKKKKK 60


RESULT 10
H12_VOLCA
ID   H12_VOLCA      STANDARD;      PRT;   240 AA.
AC   Q08865;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Histone H1-II.
GN   H1-II.
OS   Volvox carteri.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Volvocaceae; Volvox.
OX   NCBI_TaxID=3067;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=f. Nagariensis / HK10;
RX   MEDLINE=93328125; PubMed=8335260;
RA   Lindauer A., Mueller K., Schmitt R.;
RT   "Two histone H1-encoding genes of the green alga Volvox carteri with
RT   features intermediate between plant and animal genes.";
RL   Gene 129:59-68(1993).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION IS RESTRICTED TO EMBRYOGENESIS.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L07947; AAA34246.1; -.
DR   PIR; JN0748; JN0748.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR005819; Histone_H5.
DR   InterPro; IPR003216; Linkerhist_N.
DR   Pfam; PF00538; linker_histone; 1.
DR   PRINTS; PR00624; HISTONEH5.
DR   ProDom; PD000373; Linkerhist_N; 1.
DR   SMART; SM00526; H15; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW   Repeat.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN       21     98       GLOBULAR.
FT   DNA_BIND    182    185       POTENTIAL.
FT   DNA_BIND    202    205       POTENTIAL.
FT   DNA_BIND    210    213       POTENTIAL.
FT   DOMAIN      110    216       8 X 6 AA REPEATS OF P-K-K-A-[AK]-A.
FT   REPEAT      110    115       1.
FT   REPEAT      116    121       2.
FT   REPEAT      122    127       3.
FT   REPEAT      183    188       4.
FT   REPEAT      189    194       5.
FT   REPEAT      195    200       6.
FT   REPEAT      203    208       7.
FT   REPEAT      211    216       8.
SQ   SEQUENCE   240 AA;  25072 MW;  01AA389E08F421BD CRC64;

  Query Match             40.7%;  Score 85;  DB 1;  Length 240;
  Best Local Similarity   50.0%;  Pred. No. 0.54;
  Matches   23;  Conservative    7;  Mismatches   14;  Indels    2;  Gaps    1;
```

```
Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKK--KKKKKK 45
              :| |||:  || : ||| : ||| || |  ||: :||  | | :||
Db         99 KAKAAAKPKAAPKKAAAPKKAAAPKKAKAPKKEGEKKAVKPKSEKK 144

RESULT 11
PHPA_PLACH
ID   PHPA_PLACH     STANDARD;      PRT;   441 AA.
AC   Q02752;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Acidic phosphoprotein precursor (50 kDa antigen).
GN   PCEMA1.
OS   Plasmodium chabaudi.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5825;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IP-PC1;
RX   MEDLINE=93116806; PubMed=1475002;
RA   Deleersnijder W., Prasomsitti P., Tungpradubkul S., Hendrix D.,
RA   Hamers-Casterman C., Hamers R.;
RT   "Structure of a Plasmodium chabaudi acidic phosphoprotein that is
RT   associated with the host erythrocyte membrane.";
RL   Mol. Biochem. Parasitol. 56:59-68(1992).
CC   -!- FUNCTION: DURING INFECTION, THIS PHOSPHOPROTEIN PROBABLY MODULATES
CC       THE STRUCTURE OF THE RED CELL MEMBRANE TO THE ADVANTAGE OF THE
CC       PARASITE, ALTHOUGH ITS PRECISE FUNCTION IS NOT KNOWN.
CC   -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN ON THE
CC       CYTOPLASMIC FACE OF THE HOST ERYTHROCYTE MEMBRANE.
CC   -!- MISCELLANEOUS: ASSOCIATED WITH THE HOST RED CELL MEMBRANE
CC       THROUGHOUT THE ENTIRE ERYTHROCYTIC CYCLE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M95789; AAA29732.1; -.
DR   PIR; A48455; A48455.
KW   Phosphorylation; Signal; Antigen; Membrane; Repeat; Erythrocyte.
FT   SIGNAL        1     15       OR 24 (POTENTIAL).
FT   CHAIN        16    441       ACIDIC PHOSPHOPROTEIN.
FT   DOMAIN      186    313       16 X 8 AA TANDEM REPEATS.
FT   REPEAT      186    193       1-1.
FT   REPEAT      194    201       1-2.
FT   REPEAT      202    209       1-3.
FT   REPEAT      210    217       1-4.
FT   REPEAT      218    225       1-5.
FT   REPEAT      226    233       1-6.
FT   REPEAT      234    241       1-7.
FT   REPEAT      242    249       1-8.
FT   REPEAT      250    257       1-9.
FT   REPEAT      258    265       1-10.
FT   REPEAT      266    273       1-11.
FT   REPEAT      274    281       1-12.
FT   REPEAT      282    289       1-13.
FT   REPEAT      290    297       1-14.
FT   REPEAT      298    305       1-15.
FT   REPEAT      306    313       1-16.
FT   DOMAIN      353    370       2 X 9 AA TANDEM REPEATS.
FT   REPEAT      353    360       2-1.
FT   REPEAT      361    368       2-2.
FT   DOMAIN      371    417       LYS-RICH (BASIC).
FT   CARBOHYD     21     21       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   441 AA;  49708 MW;  DB85E83E795EE7E5 CRC64;
```

```
  Query Match             40.2%;  Score 84;  DB 1;  Length 441;
  Best Local Similarity   80.0%;  Pred. No. 1.1;
  Matches   16;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 KKKKKKKKKKKKKKKKKKKK 45
              ||||||:||||||:|:|||:
Db        394 KKKKKKEKKKKKKEKEKKKE 413

RESULT 12
BRD3_HUMAN
ID   BRD3_HUMAN     STANDARD;      PRT;   726 AA.
AC   Q15059; Q92645;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bromodomain-containing protein 3 (RING3-like protein).
GN   BRD3 OR RING3L OR KIAA0043.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96051398; PubMed=7584044;
RA   Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA   Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. II.
RT   The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 1:223-229(1994).
RN   [2]
RP   SEQUENCE OF 363-726 FROM N.A.
RX   MEDLINE=98038990; PubMed=9373153;
RA   Thorpe K.L., Gorman P., Thomas C., Sheer D., Trowsdale J., Beck S.;
RT   "Chromosomal localization, gene structure and transcription pattern of
RT   the ORFX gene, a homologue of the MHC-linked RING3 gene.";
RL   Gene 200:177-183(1997).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: Contains 2 bromodomains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D26362; BAA05393.1; -.
DR   EMBL; Z81330; CAB03630.1; -.
DR   HSSP; Q92831; 1B91.
DR   Genew; HGNC:1104; BRD3.
DR   MIM; 601541; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   InterPro; IPR001487; Bromodomain.
DR   Pfam; PF00439; bromodomain; 2.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 2.
DR   PROSITE; PS00633; BROMODOMAIN_1; 2.
DR   PROSITE; PS50014; BROMODOMAIN_2; 2.
KW   Bromodomain; Repeat; Nuclear protein.
FT   DOMAIN       56    115       BROMODOMAIN 1.
FT   DOMAIN      326    398       BROMODOMAIN 2.
FT   DOMAIN      487    555       LYS-RICH.
FT   DOMAIN      676    725       SER-RICH.
FT   CONFLICT    465    466       EL -> DV (IN REF. 2).
SQ   SEQUENCE   726 AA;  79541 MW;  64F526FC3C1033AA CRC64;

  Query Match             40.2%;  Score 84;  DB 1;  Length 726;
  Best Local Similarity   54.3%;  Pred. No. 1.7;
```

```
  Matches   19;  Conservative    4;  Mismatches   12;  Indels    0;  Gaps    0;

Qy         11 AAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKKK 45
                |  |  ||   |    | ||||:||:|:|||| |:|
Db        472 AVHEQLAALSQAPVNKPKKKKEKKEKEKKKKDKEK 506


RESULT 13
RL12_PYRAB
ID   RL12_PYRAB     STANDARD;      PRT;   107 AA.
AC   Q9UXS6;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   50S ribosomal protein L12P.
GN   RPL12P OR PYRAB17820 OR PAB1168.
OS   Pyrococcus abyssi.
OC   Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC   Pyrococcus.
OX   NCBI_TaxID=29292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GE5 / Orsay;
RX   PubMed=12622808;
RA   Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA   Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA   Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT   "An integrated analysis of the genome of the hyperthermophilic
RT   archaeon Pyrococcus abyssi.";
RL   Mol. Microbiol. 47:1495-1512(2003).
CC   -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC       INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC       ACCURATE TRANSLATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AJ248288; CAB50687.1; -.
DR   PIR; A75031; A75031.
DR   InterPro; IPR001813; 60s_ribosomal.
DR   Pfam; PF00428; 60s_ribosomal; 1.
KW   Ribosomal protein; Complete proteome.
SQ   SEQUENCE   107 AA;  11300 MW;  CD4239930BAAD4ED CRC64;

  Query Match             39.7%;  Score 83;  DB 1;  Length 107;
  Best Local Similarity   44.2%;  Pred. No. 0.39;
  Matches   19;  Conservative   12;  Mismatches   12;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 44
                | ||    | ||| ||| ||| |   ||:::||:::|:::  :::
Db         55 EKAAMPVAVAAAPAAAPAEAPAEEKKEEEKKEEEKEEEVSEEE 97


RESULT 14
YD33_YEAST
ID   YD33_YEAST     STANDARD;      PRT;   320 AA.
AC   Q12117;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 36.2 kDa protein in RAD28-LYS14 intergenic region.
GN   YDR033W OR YD9673.03.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Arnold W., Becker A., Jaeger W., Kuester H., Nussbaumer B.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEAL OPSIN FAMILY. HSP30
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z74329; CAA98855.1; -.
DR   EMBL; Z68196; CAA92370.1; -.
DR   PIR; S61586; S61586.
DR   SGD; S0002440; YDR033W.
DR   GO; GO:0005886; C:plasma membrane; IDA.
DR   InterPro; IPR001425; Bac_rhodopsin.
DR   Pfam; PF01036; Bac_rhodopsin; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     35     55       POTENTIAL.
FT   TRANSMEM     63     83       POTENTIAL.
FT   TRANSMEM    117    137       POTENTIAL.
FT   TRANSMEM    142    162       POTENTIAL.
FT   TRANSMEM    168    188       POTENTIAL.
FT   TRANSMEM    205    225       POTENTIAL.
FT   TRANSMEM    239    259       POTENTIAL.
FT   DOMAIN      300    318       LYS-RICH.
SQ   SEQUENCE   320 AA;  36190 MW;  4311F64EA6AA209F CRC64;

  Query Match             39.7%;  Score 83;  DB 1;  Length 320;
  Best Local Similarity   48.8%;  Pred. No. 1;
  Matches   21;  Conservative    4;  Mismatches   14;  Indels    4;  Gaps    1;

Qy          7 AEAAAAAEAAAAAEAAAA----AKKKKKKKKKKKKKKKKKKK 45
                |||  | :|  |:   ||     :| |||| || || || || :
Db        277 AEAVMAPKAPVASPRPAATPNLSKDKKKKSKKSKKSKKSKKSE 319


RESULT 15
TOLA_ECOLI
ID   TOLA_ECOLI     STANDARD;      PRT;   421 AA.
AC   P19934;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   TolA protein.
GN   TOLA OR CIM OR EXCC OR LKY OR B0739.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JM105;
RX   MEDLINE=90078104; PubMed=2687247;
RA   Levengood S.K., Webster R.E.;
RT   "Nucleotide sequences of the tolA and tolB genes and localization of
RT   their products, components of a multistep translocation system in
RT   Escherichia coli.";
RL   J. Bacteriol. 171:6600-6609(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
```

```
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   DOMAINS.
RX   MEDLINE=91296736; PubMed=2068069;
RA   Levengood S.K., Beyer W.F. Jr., Webster R.E.;
RT   "TolA: a membrane protein involved in colicin uptake contains an
RT   extended helical region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
RN   [5]
RP   INTERACTION WITH PORINS.
RX   MEDLINE=97133271; PubMed=8978668;
RA   Derouiche R., Gavioli M., Benedetti H., Prilipov A., Lazdunski C.,
RA   Lloubes R.;
RT   "TolA central domain interacts with Escherichia coli porins.";
RL   EMBO J. 15:6408-6415(1996).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX   MEDLINE=99332679; PubMed=10404600;
RA   Lubkowski J., Hennecke F., Plueckthun A., Wlodawer A.;
RT   "Filamentous phage infection: crystal structure of g3p in complex
RT   with its coreceptor, the C-terminal domain of TolA.";
RL   Structure 7:711-722(1999).
CC   -!- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
CC       COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
CC       COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
CC       BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION
CC       OF BACTERIOPHAGE DNA.
CC   -!- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPC, OMPC, PHOE
CC       AND LAMB.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M28232; AAA24683.1; -.
DR   EMBL; AE000177; AAC73833.1; -.
DR   EMBL; D90713; BAA35405.1; -.
DR   PIR; JV0057; JV0057.
DR   PDB; 1TOL; 20-MAY-99.
DR   EcoGene; EG11007; tolA.
KW   Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW   Repeat; Inner membrane; 3D-structure; Complete proteome.
FT   DOMAIN        1     13       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     14     34       POTENTIAL.
FT   DOMAIN       35    421       PERIPLASMIC (POTENTIAL).
FT   DOMAIN       48    310       DOMAIN II (ALPHA-HELICAL).
FT   DOMAIN      311    421       DOMAIN III (FUNCTIONAL).
FT   DOMAIN      224    278       10 X TANDEM REPEATS OF [ED]-K(1,2)-
FT                                A(2,4).
FT   HELIX       335    349
FT   TURN        350    351
FT   TURN        353    354
FT   HELIX       355    358
FT   TURN        359    360
FT   STRAND      363    369
FT   TURN        371    372
FT   STRAND      375    383
FT   HELIX       385    397
FT   HELIX       406    412
FT   TURN        413    414
FT   STRAND      416    421
SQ   SEQUENCE   421 AA;  43156 MW;  8B2F52B4B97C655E CRC64;

  Query Match             39.7%;  Score 83;  DB 1;  Length 421;
  Best Local Similarity   47.1%;  Pred. No. 1.3;
  Matches   24;  Conservative    5;  Mismatches   14;  Indels    8;  Gaps    1;

Qy          3 AAAAAEAAAAAEAAA--------AAEAAAAAKKKKKKKKKKKKKKKKKK 45
              | |||||  ||·||        |||||||  :||   :  :| | : :||
Db        176 AKAAAEAQKKAEAAAAALKKKAEAAEAAAAEARKKAATEAAEKAKAEAEKK 226

Search completed: January 30, 2004, 00:20:46
Job time : 11.1408 secs
```

```
OM protein - protein search, using sw model

Run on:           January 29, 2004, 23:58:52 ; Search time 46.4789 Seconds
                                               (without alignments)
                                               249.842 Million cell updates/sec

Title:            US-09-461-684C-4
Perfect score:    209
Sequence:         1 CEAAAAAEAAAAAEAAAAAE..........KKKKKKKKKKKKKKKKKKK 45

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:       830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	118	56.5	1038	5	Q8MQW9	Q8mqw9 drosophila
2	118	56.5	2347	5	Q8INH9	Q8inh9 drosophila
3	118	56.5	2451	5	Q9VG05	Q9vg05 drosophila
4	107	51.2	791	5	Q8T2U7	Q8t2u7 dictyosteli
5	105	50.2	129	11	O35807	O35807 rattus norv
6	105	50.2	720	4	Q9H6Q7	Q9h6q7 homo sapien
7	104	49.8	658	11	Q8CGI8	Q8cgi8 mus musculu
8	103	49.3	113	10	Q8LQP6	Q8lqp6 oryza sativ
9	102	48.8	467	10	Q9LL82	Q9ll82 euglena gra
10	102	48.8	517	10	Q9LXR2	Q9lxr2 arabidopsis
11	101.5	48.6	354	3	O14347	O14347 schizosacch
12	101	48.3	531	6	Q95LV6	Q95lv6 macaca fasc
13	100.5	48.1	80	10	Q8S7D3	Q8s7d3 oryza sativ
14	100	47.8	55	4	Q8N6F0	Q8n6f0 homo sapien
15	100	47.8	128	3	Q9P529	Q9p529 neurospora
16	100	47.8	168	4	Q9H5V6	Q9h5v6 homo sapien
17	100	47.8	206	5	Q8I247	Q8i247 plasmodium
18	100	47.8	215	11	Q64075	Q64075 rattus sp.
19	100	47.8	260	10	Q9LGZ9	Q9lgz9 arabidopsis
20	100	47.8	380	4	Q9NT34	Q9nt34 homo sapien
21	100	47.8	515	5	Q8SWR7	Q8swr7 drosophila
22	99	47.4	686	4	Q9NXF0	Q9nxf0 homo sapien
23	97	46.4	476	10	Q8RVV4	Q8rvv4 lycopersico
24	97	46.4	667	4	Q9HC48	Q9hc48 homo sapien
25	95	45.5	257	4	Q9H5Y3	Q9h5y3 homo sapien
26	95	45.5	489	5	Q8T3H4	Q8t3h4 drosophila
27	93	44.5	332	4	Q8IZU1	Q8izu1 homo sapien
28	92.5	44.3	372	2	Q9WWX1	Q9wwx1 pseudomonas
29	92	44.0	351	16	Q8FQY8	Q8fqy8 corynebacte
30	92	44.0	4433	5	Q8IJI5	Q8iji5 plasmodium
31	91	43.5	229	11	Q9JM93	Q9jm93 mus musculu
32	91	43.5	263	4	Q9BU76	Q9bu76 homo sapien
33	91	43.5	263	4	Q8IZH3	Q8izh3 homo sapien
34	91	43.5	556	10	Q8S405	Q8s405 hordeum vul
35	91	43.5	3351	5	Q8IBK4	Q8ibk4 plasmodium
36	90.5	43.3	368	10	Q8LCH0	Q8lch0 arabidopsis
37	90.5	43.3	368	10	Q944L9	Q944l9 arabidopsis
38	90.5	43.3	383	10	Q9LMT3	Q9lmt3 arabidopsis
39	90	43.1	39	5	Q8MSJ8	Q8msj8 drosophila
40	90	43.1	166	5	Q99281	Q99281 parechinus
41	90	43.1	180	5	Q25636	Q25636 parechinus
42	90	43.1	529	4	Q9NPX4	Q9npx4 homo sapien
43	90	43.1	628	4	Q9H5M5	Q9h5m5 homo sapien
44	90	43.1	749	5	Q967D9	Q967d9 drosophila
45	90	43.1	1062	5	Q960C4	Q960c4 drosophila

ALIGNMENTS

```
RESULT 1
Q8MQW9
ID   Q8MQW9     PRELIMINARY;     PRT;  1038 AA.
AC   Q8MQW9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   SD05989p (Fragment).
GN   CG7518.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY122252; AAM52764.1; -.
DR   FlyBase; FBgn0038108; CG7518.
DR   InterPro; IPR000104; Antifreeze_1.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR00308; ANTIFREEZEI.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER       1      1
SQ   SEQUENCE   1038 AA;  109059 MW;  80C935A2C6D8A276 CRC64;

  Query Match             56.5%;  Score 118;  DB 5;  Length 1038;
  Best Local Similarity   75.7%;  Pred. No. 0.01;
  Matches   28;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKK 39
              ||||||||||||||||| | |||::| | | ||: ||
Db        112 AAAAAEAAAAAEAAAAAAAEAAAEQKAKLKNKKQAKK 148
```

```
RESULT 2
Q8INH9
ID   Q8INH9      PRELIMINARY;      PRT;  2347 AA.
AC   Q8INH9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG7518-PB.
GN   CG7518.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003698; AAN14338.1; -.
SQ   SEQUENCE   2347 AA;  257013 MW;  23BF5FC5FFCAEA64 CRC64;

  Query Match             56.5%;  Score 118;  DB 5;  Length 2347;
  Best Local Similarity   75.7%;  Pred. No. 0.021;
  Matches   28;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAAEAAAAAKKKKKKKKKKKKK 39
              ||||||||||||||||| | |||::| | | ||: ||
Db       1374 AAAAAEAAAAAEAAAAAAAEAAAEQKAKLKNKKQAKK 1410


RESULT 3
Q9VG05
ID   Q9VG05      PRELIMINARY;      PRT;  2451 AA.
AC   Q9VG05;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CG7518 protein.
GN   CG7518.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003698; AAF54888.2; -.
DR   FlyBase; FBgn0038108; CG7518.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   PROSITE; PS00037; MYB_1; 1.
SQ   SEQUENCE   2451 AA;  266959 MW;  088A2293F27481E2 CRC64;

  Query Match             56.5%;  Score 118;  DB 5;  Length 2451;
  Best Local Similarity   75.7%;  Pred. No. 0.022;
  Matches   28;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKK 39
                |||||||||||||||| | |||::| | | ||: ||
Db       1374 AAAAAEAAAAAEAAAAAAAEAAAEQKAKLKNKKQAKK 1410

RESULT 4
Q8T2U7
ID   Q8T2U7      PRELIMINARY;      PRT;   791 AA.
AC   Q8T2U7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 92.4 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115574; AAL92183.1; -.
DR   InterPro; IPR005033; YEATS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF03366; YEATS; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   791 AA;  92375 MW;  D66CCB6DEC92352C CRC64;

  Query Match             51.2%;  Score 107;  DB 5;  Length 791;
  Best Local Similarity   56.8%;  Pred. No. 0.07;
  Matches   25;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKKK 45
                |     |     |  |  |     ||||||||||||||||||
Db        746 EKEIETEMEIGKEIEAEIEIEKEKKKKKKKKKKKKKKKKKKKKK 789

RESULT 5
O35807
ID   O35807      PRELIMINARY;      PRT;   129 AA.
AC   O35807;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   MICROVASCULAR endothelial differentiation protein 2.
GN   MDG2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Epididymis;
RX   MEDLINE=98172708; PubMed=9511718;
RA   Proels F., Loser B., Marx M.;
RT   "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA
RT   species, during in vitro angiogenesis.";
RL   Exp. Cell Res. 239:1-10(1998).
DR   EMBL; Y08769; CAA70022.1; -.
DR   InterPro; IPR000719; Prot_kinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   129 AA;  15080 MW;  38102272BBE2EDB4 CRC64;

  Query Match             50.2%;  Score 105;  DB 11;  Length 129;
  Best Local Similarity   95.5%;  Pred. No. 0.02;
  Matches   21;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         24 AAKKKKKKKKKKKKKKKKKKKK 45
              |:||||||||||||||||||||
Db         83 ASKKKKKKKKKKKKKKKKKKKK 104

RESULT 6
Q9H6Q7
ID   Q9H6Q7      PRELIMINARY;      PRT;   720 AA.
AC   Q9H6Q7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ21979 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK025632; BAB15196.1; -.
KW   Hypothetical protein.
FT   NON_TER     720    720
SQ   SEQUENCE   720 AA;  84029 MW;  A86586FEAA953D0B CRC64;

  Query Match             50.2%;  Score 105;  DB 4;  Length 720;
  Best Local Similarity   61.8%;  Pred. No. 0.095;
  Matches   21;  Conservative    6;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         12 AAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKKKK 45
              | ::   ::   : :|||||||||||||||||||
Db        678 AKKSITNSDIVSISKKKKKKKKKKKKKKKKKKKK 711

RESULT 7
Q8CGI8
ID   Q8CGI8      PRELIMINARY;      PRT;   658 AA.
AC   Q8CGI8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC035210; AAH35210.1; -.
KW   Hypothetical protein.
FT   NON_TER     658    658
SQ   SEQUENCE   658 AA;  73538 MW;  270E821463535FF9 CRC64;

  Query Match             49.8%;  Score 104;  DB 11;  Length 658;
  Best Local Similarity   64.7%;  Pred. No. 0.11;
  Matches   22;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy         11 AAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 44
              | :     |     |::||||||||||||||||||
Db        625 ATSTTCTATVNQASSKKKKKKKKKKKKKKKKKK 658


RESULT 8
Q8LQP6
ID   Q8LQP6      PRELIMINARY;      PRT;   113 AA.
AC   Q8LQP6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   OJ1117_G01.13 protein.
GN   OJ1117_G01.13.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:OJ1117_G01.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003374; BAB93330.1; -.
DR   Gramene; Q8LQP6; -.
SQ   SEQUENCE   113 AA;  13660 MW;  597DB0EDEB2AA3EF CRC64;

  Query Match             49.3%;  Score 103;  DB 10;  Length 113;
  Best Local Similarity   73.3%;  Pred. No. 0.027;
  Matches   22;  Conservative    1;  Mismatches    7;  Indels    0;  Gaps    0;

Qy         16 AAAAEAAAAAKKKKKKKKKKKKKKKKKKKKKK 45
              | : |       ||||||||||||||||||
Db          3 ATSLEHIKKKKKKKKKKKKKKKKKKKKKKK 32


RESULT 9
Q9LL82
ID   Q9LL82      PRELIMINARY;      PRT;   467 AA.
AC   Q9LL82;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Cbf5p.
OS   Euglena gracilis.
OC   Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
OX   NCBI_TaxID=3039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=streptomycin-bleached strain;
RX   MEDLINE=20330353; PubMed=10871366;
RA   Watanabe Y., Gray M.W.;
RT   "Evolutionary appearance of genes encoding proteins associated with
RT   box H/ACA snoRNAs: Cbf5p in Euglena gracilis, an early diverging
RT   eukaryote, and candidate Gar1p and Nop10p homologs in
RT   archaebacteria.";
RL   Nucleic Acids Res. 28:2342-2352(2000).
DR   EMBL; AF234319; AAF77119.1; -.
DR   InterPro; IPR004802; Cbf5.
DR   InterPro; IPR002478; PUA.
DR   InterPro; IPR002501; TruB_N.
DR   InterPro; IPR004521; Unchar_dom_2.
DR   Pfam; PF01472; PUA; 1.
DR   Pfam; PF01509; TruB_N; 1.
DR   SMART; SM00359; PUA; 1.
DR   TIGRFAMs; TIGR00425; CBF5; 1.
DR   TIGRFAMs; TIGR00451; unchar_dom_2; 1.
SQ   SEQUENCE   467 AA;  52385 MW;  401089B66507BA7B CRC64;

  Query Match             48.8%;  Score 102;  DB 10;  Length 467;
  Best Local Similarity   55.8%;  Pred. No. 0.12;
  Matches   24;  Conservative    5;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 EAAAAAEAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 44
              |  | ||||    |   || |      ||| ||:||:||:||:||:|
Db        420 EELAYAEAAKKREREAAGEDEKDAKKAKKEKKEKKEKKEKKEK 462


RESULT 10
Q9LXR2
ID   Q9LXR2      PRELIMINARY;      PRT;   517 AA.
AC   Q9LXR2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 59.7 kDa protein.
GN   T20N10_250.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA   Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353032; CAB88307.1; -.
DR   InterPro; IPR001810; F-box.
DR   InterPro; IPR006566; FBD.
DR   Pfam; PF00646; F-box; 1.
DR   SMART; SM00579; FBD; 1.
DR   SMART; SM00256; FBOX; 1.
DR   PROSITE; PS50181; FBOX; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   517 AA;  59689 MW;  EC6D957D01F86E70 CRC64;

  Query Match             48.8%;  Score 102;  DB 10;  Length 517;
  Best Local Similarity   80.8%;  Pred. No. 0.13;
  Matches   21;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         20 EAAAAAKKKKKKKKKKKKKKKKKKKK 45
              |     ||||||||||||||||||||
Db        440 ERVGKKKKKKKKKKKKKKKKKKKKKK 465


RESULT 11
O14347
ID   O14347      PRELIMINARY;      PRT;   354 AA.
AC   O14347;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Zuotin-like protein (Fragment).
GN   SPBC30D10.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Beck A., Reinhardt R., Lyne M., Rajandream M.A., Barrell B.G.;
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z97992; CAB10796.1; -.
DR   HSSP; P25685; 1HDJ.
DR   InterPro; IPR001623; DnaJ_N.
DR   Pfam; PF00226; DnaJ; 1.
DR   SMART; SM00271; DnaJ; 1.
DR   PROSITE; PS00636; DNAJ_1; 1.
DR   PROSITE; PS50076; DNAJ_2; 1.
KW   DNA-binding.
FT   NON_TER     354    354
SQ   SEQUENCE   354 AA;  40290 MW;  6071B58A3B60F558 CRC64;

  Query Match             48.6%;  Score 101.5;  DB 3;  Length 354;
  Best Local Similarity   56.0%;  Pred. No. 0.1;
  Matches   28;  Conservative    3;  Mismatches   12;  Indels    7;  Gaps    1;

Qy          2 EAAAAA-------EAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 44
              ||||||       |  || |||| | |||| || |: ||| :|: ||  |
Db        221 EAAAAAQKKKEEEERRAAEEAAAKASAAAANKKAKEDKKKAQKRDKKVVK 270


RESULT 12
Q95LV6
ID   Q95LV6      PRELIMINARY;      PRT;   531 AA.
AC   Q95LV6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 61.4 kDa protein (Fragment).
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA   Terao K., Sugano S.;
RT   "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT   libraries.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB071085; BAB64479.1; -.
KW   Hypothetical protein.
FT   NON_TER     531    531
SQ   SEQUENCE   531 AA;  61389 MW;  B55996B4F5CDD60C CRC64;

  Query Match             48.3%;  Score 101;  DB 6;  Length 531;
  Best Local Similarity   95.2%;  Pred. No. 0.16;
  Matches   20;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         25 AKKKKKKKKKKKKKKKKKKKK 45
              :||||||||||||||||||||
Db        501 SKKKKKKKKKKKKKKKKKKKK 521


RESULT 13
Q8S7D3
ID   Q8S7D3      PRELIMINARY;      PRT;    80 AA.
AC   Q8S7D3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 9.4 kDa protein.
GN   OSJNBA0057L21.23.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA   Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA   Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA   VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA   Salzberg S.L., White O., Fraser C.M.;
RT   "Oryza sativa chromosome 10 BAC OSJNBa0057L21 genomic sequence.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC087599; AAL79706.1; -.
DR   Gramene; Q8S7D3; -.
KW   Hypothetical protein.
SQ   SEQUENCE   80 AA;  9362 MW;  0177C863133B21D8 CRC64;

  Query Match             48.1%;  Score 100.5;  DB 10;  Length 80;
  Best Local Similarity   49.1%;  Pred. No. 0.032;
  Matches   27;  Conservative    1;  Mismatches   16;  Indels   11;  Gaps    1;

Qy          2 EAAAAA-----------EAAAAAEAAAAAEAAAAAKKKKKKKKKKKKKKKKKK 45
              |||  |           | ||     :      ||||||||||||||||||
Db         15 EAAGGATPGGRWGHWWWPATAATRVFLSVIHLDKKKKKKKKKKKKKKKKKKKK 69


RESULT 14
Q8N6F0
ID   Q8N6F0      PRELIMINARY;      PRT;    55 AA.
AC   Q8N6F0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to LOC201361.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC030525; AAH30525.1; -.
SQ   SEQUENCE   55 AA;  7251 MW;  0906032B284006BA CRC64;

  Query Match             47.8%;  Score 100;  DB 4;  Length 55;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 KKKKKKKKKKKKKKKKKKKK 45
              ||||||||||||||||||||
Db         21 KKKKKKKKKKKKKKKKKKKK 40


RESULT 15
Q9P529
ID   Q9P529      PRELIMINARY;      PRT;   128 AA.
AC   Q9P529;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 15.2 kDa protein.
GN   B24H17.160.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
```

```
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356815; CAB92638.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   128 AA;  15157 MW;  8C7C65C3DFB70765 CRC64;

  Query Match             47.8%;  Score 100;  DB 3;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 0.055;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         26 KKKKKKKKKKKKKKKKKKKK 45
              ||||||||||||||||||||
Db         71 KKKKKKKKKKKKKKKKKKKK 90


Search completed: January 30, 2004, 00:24:41
Job time : 46.4789 secs
```

```
OM protein - protein search, using sw model

Run on:         January 30, 2004, 00:13:12 ; Search time 18.7981 Seconds
                                             (without alignments)
                                             225.098 Million cell updates/sec

Title:          US-09-461-684C-5
Perfect score:  243
Sequence:       1 CGLFGAIAGFIENGWEGMID..........KKKKKKKKKKKKKKKKKKK 44

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_76:*
                1:  pir1:*
                2:  pir2:*
                3:  pir3:*
                4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	134	55.1	550	1	HMIVS2	hemagglutinin prec
2	134	55.1	550	1	HMIVS3	hemagglutinin prec
3	134	55.1	550	1	HMIV77	hemagglutinin prec
4	134	55.1	550	1	HMIV80	hemagglutinin prec
5	134	55.1	550	1	HMIV33	hemagglutinin prec
6	134	55.1	550	1	HMIV89	hemagglutinin prec
7	134	55.1	550	1	HMIV21	hemagglutinin prec
8	134	55.1	550	1	HMIV98	hemagglutinin prec
9	134	55.1	550	1	HMIV15	hemagglutinin prec
10	134	55.1	550	2	JQ1153	hemagglutinin prec
11	134	55.1	550	2	JQ1154	hemagglutinin prec
12	134	55.1	550	2	JQ1155	hemagglutinin prec
13	134	55.1	566	1	HMIVH	hemagglutinin prec
14	134	55.1	566	1	HMIVHA	hemagglutinin prec
15	134	55.1	566	1	HMIVHM	hemagglutinin prec
16	134	55.1	566	1	HMIVDU	hemagglutinin prec
17	134	55.1	570	1	A45591	hemagglutinin prec
18	134	55.1	570	2	S22013	hemagglutinin prec
19	133	54.7	561	1	HMIV49	hemagglutinin prec
20	133	54.7	561	1	HMIV84	hemagglutinin prec
21	132	54.3	565	1	HMIVE1	hemagglutinin prec
22	132	54.3	565	1	HMIVE3	hemagglutinin prec
23	132	54.3	566	1	HMIV6	hemagglutinin prec
24	132	54.3	567	1	HMIVV	hemagglutinin prec
25	131	53.9	362	2	S38637	hemagglutinin - in
26	131	53.9	550	1	HMIV86	hemagglutinin prec
27	131	53.9	560	1	HMIVT7	hemagglutinin prec
28	131	53.9	565	1	HMIVE2	hemagglutinin prec
29	131	53.9	565	1	HMIVE4	hemagglutinin prec
30	131	53.9	565	1	HMIVE5	hemagglutinin prec
31	131	53.9	565	1	HMIVE6	hemagglutinin prec
32	131	53.9	565	1	HMIVE7	hemagglutinin prec
33	131	53.9	565	1	HMIVE8	hemagglutinin prec
34	131	53.9	565	1	HMIVE9	hemagglutinin prec
35	131	53.9	565	1	HMIVET	hemagglutinin prec
36	131	53.9	565	1	HMIVEE	hemagglutinin prec
37	131	53.9	565	2	S33703	hemagglutinin - in
38	131	53.9	570	2	S22014	hemagglutinin prec
39	131	53.9	570	2	S22015	hemagglutinin prec
40	131	53.9	570	2	S22016	hemagglutinin prec
41	131	53.9	570	2	S22017	hemagglutinin prec
42	131	53.9	570	2	S22018	hemagglutinin prec
43	131	53.9	570	2	S22020	hemagglutinin prec
44	131	53.9	570	2	S22021	hemagglutinin prec
45	131	53.9	570	2	S22029	hemagglutinin prec

ALIGNMENTS

```
RESULT 1
HMIVS2
hemagglutinin precursor - influenza A virus (strain A/swine/126/82) (fragment)
C;Species: influenza A virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C;Accession: A29971
R;Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A;Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A;Reference number: A94370; MUID:88101364; PMID:3336940
A;Accession: A29971
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M19056; NID:g324208
A;Note: the sequence in GenBank entry FLAHAPA, release 106, (PID:g324209) differs from tl
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match          55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 2
HMIVS3
hemagglutinin precursor - influenza A virus (strain A/swine/81/78) (fragment)
C;Species: influenza A virus
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 18-Sep-1998
C;Accession: B29971
R;Kida, H.; Shortridge, K.F.; Webster, R.G.
Virology 162, 160-166, 1988
A;Title: Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China.
A;Reference number: A94370; MUID:88101364; PMID:3336940
A;Accession: B29971
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M19057; NID:g324210
A;Note: the sequence in GenBank entry FLAHAPB, release 106, (PID:g324211) differs from tl
C;Genetics:
A;Map position: segment 4
```

```
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;300-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 3
HMIV77
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/5/77) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: A27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: A27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16737; NID:g324081; PIDN:AAA43143.1; PID:g324082
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 4
HMIV80
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/8/80) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: B27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: B27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16738; NID:g324083
A;Note: the translation in Fig. 2 is inconsistent with the nucleotide sequence in Fig. 1
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 5
HMIV33
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/33/80) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: C27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: C27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16739; NID:g324085; PIDN:AAA43145.1; PID:g324086
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 6
HMIV89
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/7/82) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: D27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: D27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16740; NID:g324087; PIDN:AAA43146.1; PID:g324088
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
```

F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

```
  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352
```

RESULT 7
HMIV21
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/21/82) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: E27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: E27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16741; NID:g324089
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;7,8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted

```
  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352
```

RESULT 8
HMIV98
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/9/85) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Sep-1998
C;Accession: F27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: F27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16742; NID:g324091
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

```
  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352
```

RESULT 9
HMIV15
hemagglutinin precursor - influenza A virus (strain A/duck/Hokkaido/10/85) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: G27813
R;Kida, H.; Kawaoka, Y.; Naeve, C.W.; Webster, R.G.
Virology 159, 109-119, 1987
A;Title: Antigenic and genetic conservation of H3 influenza virus in wild ducks.
A;Reference number: A94363; MUID:87265458; PMID:2440178
A;Accession: G27813
A;Molecule type: genomic RNA
A;Residues: 1-550 <KID>
A;Cross-references: GB:M16743; NID:g324093; PIDN:AAA43149.1; PID:g324094
C;Genetics:
A;Map position: segment 4
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-550/Product: hemagglutinin HA2 #status predicted <HA2>
F;520-536/Domain: transmembrane #status predicted <TM1>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;14-466,52-277,64-76,139-473,281-305/Disulfide bonds: #status predicted
F;539,546,549/Binding site: palmitate (Cys) (covalent) #status predicted

```
  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352
```

RESULT 10
JQ1153
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/7/75) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1153
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1153
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00929; NID:g221279; PIDN:BAA00769.1; PID:g221280
A;Note: the authors translated the codon GGG for residue 218 as Glu
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
  Query Match             55.1%;  Score 134;  DB 2;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
```

```
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 11
JQ1154
hemagglutinin precursor - influenza A virus (strain A/goose/Hong Kong/10/76) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1154
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1154
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00930; NID:g221273; PIDN:BAA00770.1; PID:g221274
A;Note: the authors translated the codon GGG for residue 218 as Glu
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 2;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 12
JQ1155
hemagglutinin precursor - influenza A virus (strain A/duck/Hong Kong/64/76) (fragment)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000
C;Accession: JQ1155
R;Yasuda, J.; Shortridge, K.F.; Shimizu, Y.; Kida, H.
J. Gen. Virol. 72, 2007-2010, 1991
A;Title: Molecular evidence for a role of domestic ducks in the introduction of avian H3
A;Reference number: JQ1153; MUID:91341491; PMID:1875195
A;Accession: JQ1155
A;Molecule type: genomic RNA
A;Residues: 1-550 <YAS>
A;Cross-references: GB:D00931; NID:g221277; PIDN:BAA00771.1; PID:g221278
A;Note: the authors translated the codon GGG for residue 218 as Glu, GCC for residue 538
A;Note: residues 528-532 are not shown in this publication
C;Superfamily: influenza virus hemagglutinin
C;Keywords: glycoprotein; homotrimer
F;1-328/Product: hemagglutinin HA1 #status predicted <HA1>
F;330-545/Product: hemagglutinin HA2 #status predicted <HA2>
F;8,22,38,165,285,483/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 2;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 13
HMIVH
hemagglutinin precursor - influenza A virus
C;Species: influenza A virus
C;Date: 28-Feb-1981 #sequence_revision 28-Feb-1981 #text_change 22-Oct-1999
C;Accession: A93705; A93233; A04051; A93231; A94441
R;Both, G.W.; Sleigh, M.J.
Nucleic Acids Res. 8, 2561-2575, 1980
A;Title: Complete nucleotide sequence of the haemagglutinin gene from a human influenza v
A;Reference number: A93705; MUID:81053698; PMID:6253883
A;Accession: A93705
A;Molecule type: genomic RNA
A;Residues: 1-566 <BOT>
A;Cross-references: GB:V01103
A;Experimental source: strain A/NT/60/68/29C
A;Note: human influenza strain A/NT/60/68/29C is a laboratory-isolated variant of A/NT/6(
R;Dopheide, T.A.; Ward, C.W.
FEBS Lett. 110, 181-183, 1980
A;Title: The disulphide bonds of a Hong Kong influenza virus hemagglutinin.
A;Reference number: A91276; MUID:80179105; PMID:6768586
A;Contents: annotation; disulfide bonds
R;Gething, M.J.; Bye, J.; Skehel, J.; Waterfield, M.
Nature 287, 301-306, 1980
A;Title: Cloning and DNA sequence of double-stranded copies of haemagglutinin genes from
A;Reference number: A93233; MUID:81030852; PMID:7421990
A;Accession: A93233
A;Molecule type: genomic RNA
A;Residues: 1-24,'S',26,'D',28-159,'G',161-197,'I',199-241,'L',243-249 <GET>
A;Experimental source: strain X-31[H3]
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;536-552/Domain: transmembrane #status predicted <TM1>
F;30-482,68-293,80-92,155-489,297-321/Disulfide bonds: #status experimental
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 14
HMIVHA
hemagglutinin precursor - influenza A virus (strain A/Aichi/2/68)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
C;Accession: A93231; A04051
R;Verhoeyen, M.; Fang, R.; Min Jou, W.; Devos, R.; Huylebroeck, D.; Saman, E.; Fiers, W.
Nature 286, 771-776, 1980
A;Title: Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/
A;Reference number: A93231; MUID:80254693; PMID:7402351
A;Accession: A93231
A;Molecule type: genomic RNA
A;Residues: 1-566 <VER>
A;Cross-references: GB:J02090; NID:g324131; PIDN:AAA43178.1; PID:g324132
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 15
```

```
HMIVHM
hemagglutinin precursor - influenza A virus (strain A/Mem/102/72)
N;Contains: hemagglutinin HA1; hemagglutinin HA2
C;Species: influenza A virus
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 31-Mar-2000
C;Accession: A94441; A04051
R;Sleigh, M.J.; Both, G.W.; Brownlee, G.G.; Bender, V.J.; Moss, B.A.
in Structure and Variation in Influenza Virus, Laver, G., and Air, G., eds., pp.69-79, E
A;Title: The haemagglutinin gene of influenza A virus: nucleotide sequence analysis of c
A;Reference number: A94441
A;Accession: A94441
A;Molecule type: genomic RNA
A;Residues: 1-566 <SLE>
C;Superfamily: influenza virus hemagglutinin
C;Keywords: hemagglutinin; homotrimer; lipoprotein; thiolester bond
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-344/Product: hemagglutinin HA1 #status predicted <HA1>
F;346-566/Product: hemagglutinin HA2 #status predicted <HA2>
F;555,562,565/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             55.1%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-06;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


Search completed: January 30, 2004, 00:26:22
Job time : 18.7981 secs
```

```
OM protein - protein search, using sw model

Run on:         January 29, 2004, 23:58:27 ; Search time 9.91549 Seconds
                                             (without alignments)
                                             208.681 Million cell updates/sec

Title:          US-09-461-684C-5
Perfect score:  243
Sequence:       1 CGLFGAIAGFIENGWEGMID..........KKKKKKKKKKKKKKKKKKK 44

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:      127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_41:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	134	55.1	550	1	HEMA_IADH1	P12582 influenza a
2	134	55.1	550	1	HEMA_IADH2	P12583 influenza a
3	134	55.1	550	1	HEMA_IADH3	P12584 influenza a
4	134	55.1	550	1	HEMA_IADH4	P12585 influenza a
5	134	55.1	550	1	HEMA_IADH5	P12586 influenza a
6	134	55.1	550	1	HEMA_IADH6	P12587 influenza a
7	134	55.1	550	1	HEMA_IADH7	P12588 influenza a
8	134	55.1	550	1	HEMA_IADHK	P43257 influenza a
9	134	55.1	550	1	HEMA_IADHL	P43258 influenza a
10	134	55.1	550	1	HEMA_IAGHK	P43260 influenza a
11	134	55.1	550	1	HEMA_IAZH2	P11133 influenza a
12	134	55.1	550	1	HEMA_IAZH3	P11134 influenza a
13	134	55.1	566	1	HEMA_IAAIC	P03437 influenza a
14	134	55.1	566	1	HEMA_IADA3	P26134 influenza a
15	134	55.1	566	1	HEMA_IADU3	P03442 influenza a
16	134	55.1	566	1	HEMA_IAMAO	P26138 influenza a
17	134	55.1	566	1	HEMA_IAME1	P03449 influenza a
18	134	55.1	566	1	HEMA_IAME2	P03439 influenza a
19	134	55.1	566	1	HEMA_IANT6	P03436 influenza a
20	134	55.1	570	1	HEMA_IAHC6	P26094 influenza a
21	134	55.1	570	1	HEMA_IAHPR	P26101 influenza a
22	133	54.7	561	1	HEMA_IACKG	P12581 influenza a
23	133	54.7	561	1	HEMA_IAMIN	P12439 influenza a
24	132	54.3	565	1	HEMA_IAHTO	P17000 influenza a
25	132	54.3	565	1	HEMA_IAHUR	P17002 influenza a
26	132	54.3	566	1	HEMA_IAEN7	P03440 influenza a
27	132	54.3	566	1	HEMA_IAZCO	P26139 influenza a
28	132	54.3	567	1	HEMA_IAVI7	P03435 influenza a
29	131	53.9	550	1	HEMA_IAME6	P12589 influenza a
30	131	53.9	560	1	HEMA_IATKR	P03458 influenza a
31	131	53.9	565	1	HEMA_IAHAL	P16994 influenza a
32	131	53.9	565	1	HEMA_IAHFO	P16995 influenza a
33	131	53.9	565	1	HEMA_IAHK6	P19699 influenza a
34	131	53.9	565	1	HEMA_IAHK7	P16996 influenza a
35	131	53.9	565	1	HEMA_IAHMI	P15658 influenza a
36	131	53.9	565	1	HEMA_IAHNM	P16997 influenza a
37	131	53.9	565	1	HEMA_IAHRO	P16998 influenza a
38	131	53.9	565	1	HEMA_IAHSA	P16999 influenza a
39	131	53.9	565	1	HEMA_IAHSU	Q08011 influenza a
40	131	53.9	565	1	HEMA_IAHTE	P17001 influenza a
41	131	53.9	566	1	HEMA_IAZUK	P26141 influenza a
42	131	53.9	570	1	HEMA_IAHC7	P26095 influenza a
43	131	53.9	570	1	HEMA_IAHCD	P26096 influenza a
44	131	53.9	570	1	HEMA_IAHDE	P26097 influenza a
45	131	53.9	570	1	HEMA_IAHLE	P26098 influenza a

ALIGNMENTS

```
RESULT 1
HEMA_IADH1
ID   HEMA_IADH1     STANDARD;      PRT;   550 AA.
AC   P12582; Q84021; Q84022;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/5/77).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11357;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16737; AAA43143.1; -.
DR   HSSP; P03437; 3HMG.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61705 MW;  7E7ACFE716FC969A CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352

RESULT 2
HEMA_IADH2
ID   HEMA_IADH2     STANDARD;      PRT;   550 AA.
AC   P12583; Q84011;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/8/80).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11358;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16738; AAA43144.1; -.
DR   PIR; B27813; HMIV80.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    137    137       K -> N (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61659 MW;  A107023ACC9CC353 CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 3
HEMA_IADH3
ID   HEMA_IADH3     STANDARD;      PRT;   550 AA.
AC   P12584; Q84012; Q89793;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11359;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16739; AAA43145.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61577 MW;  6C30BF67CFDCB7DE CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 4
HEMA_IADH4
ID   HEMA_IADH4     STANDARD;      PRT;   550 AA.
AC   P12585; Q84013; Q84014;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/7/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11360;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16740; AAA43146.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61664 MW;  A16B2CF8CBBBD9D0 CRC64;

  Query Match            55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 5
HEMA_IADH5
ID   HEMA_IADH5     STANDARD;      PRT;   550 AA.
AC   P12586; Q84015; Q84016;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/21/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11361;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16741; AAA43147.1; -.
DR   PIR; E27813; HMIV21.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      7      7       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    178    179       YV -> VI (IN PIR DATA BANK).
FT   CONFLICT    388    388       K -> T (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61856 MW;  48401C867A15BF8C CRC64;

  Query Match            55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 6
HEMA_IADH6
ID   HEMA_IADH6     STANDARD;      PRT;   550 AA.
AC   P12587; Q84017;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11362;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M16742; AAA43148.1; -.
DR   PIR; F27813; HMIV98.
DR   HSSP; P03437; 1HGJ.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      8      8       Y -> N (IN PIR DATA BANK).
SQ   SEQUENCE   550 AA;  61711 MW;  67BCD85F44736CFE CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 7
HEMA_IADH7
ID   HEMA_IADH7     STANDARD;      PRT;   550 AA.
AC   P12588; Q84018; Q89470;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11363;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87265458; PubMed=2440178;
RA   Kida H., Kawaoka Y., Naeve C.W., Webster R.G.;
RT   "Antigenic and genetic conservation of H3 influenza virus in wild
RT   ducks.";
RL   Virology 159:109-119(1987).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M16743; AAA43149.1; -.
DR   HSSP; P03437; 3HMG.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61761 MW;  6EF81793281D53EB CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 8
HEMA_IADHK
ID   HEMA_IADHK     STANDARD;      PRT;   550 AA.
AC   P43257;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hong Kong/7/75).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11364;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D00929; BAA00769.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61549 MW;  864639B829FE1BA9 CRC64;
```

```
  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 9
HEMA_IADHL
ID   HEMA_IADHL     STANDARD;      PRT;   550 AA.
AC   P43258;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Duck/Hong Kong/64/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=45412;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D00931; BAA00771.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61718 MW;  A351C56789E4BE9A CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 10
HEMA_IAGHK
ID   HEMA_IAGHK     STANDARD;      PRT;   550 AA.
AC   P43260;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Goose/Hong Kong/10/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=45414;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91341491; PubMed=1875195;
RA   Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
RT   "Molecular evidence for a role of domestic ducks in the introduction
RT   of avian H3 influenza viruses to pigs in southern China, where the
RT   A/Hong Kong/68 (H3N2) strain emerged.";
RL   J. Gen. Virol. 72:2007-2010(1991).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D00930; BAA00770.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61676 MW;  9A1E094DA28BACD2 CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 11
HEMA_IAZH2
ID   HEMA_IAZH2     STANDARD;      PRT;   550 AA.
AC   P11133; Q84019; Q84020;
DT   01-JUL-1989 (Rel. 11, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE   chain] (Fragment).
GN   HA.
```

```
OS   Influenza A virus (strain A/Swine/Hong Kong/81/78).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11497;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88101364; PubMed=3336940;
RA   Kida H., Shortridge K.F., Webster R.G.;
RT   "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
RT   in China.";
RL   Virology 162:160-166(1988).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M19057; AAA43212.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Hemagglutinin; Envelope protein; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61437 MW;  1F2A7E758C531CE8 CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 12
HEMA_IAZH3
ID   HEMA_IAZH3     STANDARD;      PRT;   550 AA.
AC   P11134; Q84025; Q84026;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin [Contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE   chain] (Fragment).
GN   HA.
OS   Influenza A virus (strain A/Swine/Hong Kong/126/82).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11498;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88101364; PubMed=3336940;
RA   Kida H., Shortridge K.F., Webster R.G.;
RT   "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs
RT   in China.";
RL   Virology 162:160-166(1988).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M19056; AAA43211.1; ALT_TERM.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Hemagglutinin; Envelope protein; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN         1    328       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       330    550       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD      8      8       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     22     22       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    165    165       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   550 AA;  61580 MW;  991F6D8BC02F24F2 CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 550;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        330 GLFGAIAGFIENGWEGMIDGWYG 352


RESULT 13
HEMA_IAAIC
ID   HEMA_IAAIC     STANDARD;      PRT;   566 AA.
AC   P03437;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Aichi/2/68).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=150147;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=80254693; PubMed=7402351;
RA   Verhoeyen M., Fang R., Min Jou W., Devos R., Huylebroeck D.,
RA   Saman E., Fiers W.;
RT   "Antigenic drift between the haemagglutinin of the Hong Kong
RT   influenza strains A/Aichi/2/68 and A/Victoria/3/75.";
RL   Nature 286:771-776(1980).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX   MEDLINE=81123029; PubMed=7464906;
RA   Wilson I.A., Skehel J.J., Wiley D.C.;
RT   "Structure of the haemagglutinin membrane glycoprotein of influenza
RT   virus at 3-A resolution.";
RL   Nature 289:366-373(1981).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE=88232903; PubMed=3374584;
```

```
RA   Weis W.I., Brown J.H., Cusack S.C., Paulson J.C., Skehel J.J.,
RA   Wiley D.C.;
RT   "Structure of the influenza virus haemagglutinin complexed with its
RT   receptor, sialic acid.";
RL   Nature 333:426-431(1988).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY OF A MUTANT WITH GLY-457.
RX   MEDLINE=90107940; PubMed=2295311;
RA   Weis W.I., Cusack S.C., Brown J.H., Daniels R.S., Skehel J.J.,
RA   Wiley D.C.;
RT   "The structure of a membrane fusion mutant of the influenza virus
RT   haemagglutinin.";
RL   EMBO J. 9:17-24(1990).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY.
RX   MEDLINE=90230310; PubMed=2329580;
RA   Weis W.I., Bruenger A.T., Skehel J.J., Wiley D.C.;
RT   "Refinement of the influenza virus hemagglutinin by simulated
RT   annealing.";
RL   J. Mol. Biol. 212:737-761(1990).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX   MEDLINE=94352388; PubMed=8072525;
RA   Bullough P.A., Hughson F.M., Skehel J.J., Wiley D.C.;
RT   "Structure of influenza haemagglutinin at the pH of membrane fusion.";
RL   Nature 371:37-43(1994).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (3.25 ANGSTROMS).
RX   MEDLINE=98120975; PubMed=9461077;
RA   Fleury D., Wharton S.A., Skehel J.J., Knossow M., Bizebard T.;
RT   "Antigen distortion allows influenza virus to escape neutralization.";
RL   Nat. Struct. Biol. 5:119-123(1998).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J02090; AAA43178.1; -.
DR   EMBL; V01085; CAA24269.1; -.
DR   PIR; A93231; HMIVHA.
DR   PDB; 2HMG; 31-OCT-93.
DR   PDB; 3HMG; 31-OCT-93.
DR   PDB; 4HMG; 31-OCT-93.
DR   PDB; 5HMG; 31-JAN-94.
DR   PDB; 1HGD; 31-JAN-94.
DR   PDB; 1HGE; 31-JAN-94.
DR   PDB; 1HGF; 31-JAN-94.
DR   PDB; 1HGG; 31-JAN-94.
DR   PDB; 1HGH; 31-JAN-94.
DR   PDB; 1HGI; 31-JAN-94.
DR   PDB; 1HGJ; 31-JAN-94.
DR   PDB; 1HTM; 14-FEB-95.
DR   PDB; 2VIR; 29-APR-98.
DR   PDB; 2VIS; 29-APR-98.
DR   PDB; 2VIT; 29-APR-98.
DR   PDB; 2VIU; 29-APR-98.
DR   PDB; 1EO8; 29-NOV-00.
DR   PDB; 1HA0; 22-DEC-99.
DR   PDB; 1J8H; 13-MAR-02.
DR   PDB; 1KEN; 24-APR-02.
DR   PDB; 1QFU; 29-DEC-99.
DR   PDB; 1QU1; 05-JAN-00.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   DISULFID     30    482       INTERCHAIN.
FT   DISULFID     68    293
FT   DISULFID     80     92
FT   DISULFID    113    155
FT   DISULFID    297    321
FT   DISULFID    489    493
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .).
FT   STRAND       27     35
FT   STRAND       40     42
FT   STRAND       50     52
FT   STRAND       55     57
FT   STRAND       59     60
FT   STRAND       67     70
FT   STRAND       74     76
FT   TURN         78     79
FT   HELIX        82     87
FT   TURN         88     88
FT   HELIX        90     95
FT   TURN         96     97
FT   STRAND       99     99
FT   STRAND      102    105
FT   TURN        107    108
FT   STRAND      116    117
FT   TURN        119    120
FT   HELIX       121    131
FT   STRAND      133    133
FT   STRAND      136    138
FT   TURN        144    145
FT   STRAND      146    147
FT   STRAND      152    157
FT   TURN        158    159
FT   STRAND      160    162
FT   TURN        165    166
FT   STRAND      167    169
FT   STRAND      171    173
FT   TURN        174    175
FT   STRAND      176    176
FT   STRAND      180    185
FT   STRAND      192    200
FT   HELIX       204    211
FT   STRAND      217    221
FT   STRAND      226    229
FT   STRAND      239    239
FT   TURN        240    241
FT   STRAND      242    242
FT   STRAND      245    253
FT   TURN        255    256
FT   STRAND      258    265
FT   STRAND      267    270
FT   STRAND      272    275
FT   STRAND      282    285
FT   STRAND      290    294
FT   STRAND      297    299
FT   TURN        300    301
FT   STRAND      302    304
FT   STRAND      310    311
FT   STRAND      318    320
FT   STRAND      323    324
FT   STRAND      331    333
FT   STRAND      337    337
FT   TURN        347    348
```

```
FT   TURN        350    350
FT   STRAND      351    351
FT   TURN        352    354
FT   STRAND      355    355
FT   STRAND      359    360
FT   TURN        361    361
FT   STRAND      367    373
FT   TURN        374    375
FT   STRAND      376    382
FT   HELIX       383    400
FT   TURN        401    401
FT   STRAND      406    407
FT   HELIX       421    471
FT   HELIX       472    474
FT   STRAND      475    477
FT   STRAND      482    485
FT   HELIX       491    498
FT   TURN        499    500
FT   HELIX       504    515
FT   TURN        518    519
SQ   SEQUENCE   566 AA;  63415 MW;  E395659C23CAFECA CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 14
HEMA_IADA3
ID   HEMA_IADA3     STANDARD;      PRT;   566 AA.
AC   P26134;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Duck/Alberta/78/76).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11348;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92114135; PubMed=1731092;
RA   Bean W.J., Schell M., Katz J., Kawaoka Y., Naeve C., Gorman O.,
RA   Webster R.G.;
RT   "Evolution of the H3 influenza virus hemagglutinin from human and
RT   nonhuman hosts.";
RL   J. Virol. 66:1129-1138(1992).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M73771; -; NOT_ANNOTATED_CDS.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     23     23       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63534 MW;  FE19AB6FF9415B89 CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 15
HEMA_IADU3
ID   HEMA_IADU3     STANDARD;      PRT;   566 AA.
AC   P03442;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE   Hemagglutinin HA2 chain].
GN   HA.
OS   Influenza A virus (strain A/Duck/Ukraine/1/63).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=11374;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82025542; PubMed=6169439;
RA   Fang R., Min Jou W., Huylebroeck D., Devos R., Fiers W.;
RT   "Complete structure of A/duck/Ukraine/63 influenza hemagglutinin
RT   gene: animal virus as progenitor of human H3 Hong Kong 1968 influenza
RT   hemagglutinin.";
RL   Cell 25:315-323(1981).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION.
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; V01087; CAA24271.1; -.
DR   PDB; 1IBN; 08-AUG-01.
DR   PDB; 1IBO; 08-AUG-01.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Hemagglutinin; Glycoprotein; Signal; 3D-structure.
FT   SIGNAL        1     16
FT   CHAIN        17    344       HEMAGGLUTININ HA1 CHAIN.
FT   CHAIN       346    566       HEMAGGLUTININ HA2 CHAIN.
FT   CARBOHYD     24     24       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     97     97       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   566 AA;  63530 MW;  E70F87F0AE1178F4 CRC64;

  Query Match             55.1%;  Score 134;  DB 1;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


Search completed: January 30, 2004, 00:20:46
Job time : 9.91549 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          January 29, 2004, 23:58:52 ; Search time 45.446 Seconds
                                           (without alignments)
                                           249.842 Million cell updates/sec

Title:           US-09-461-684C-5
Perfect score:   243
Sequence:        1 CGLFGAIAGFIENGWEGMID..........KKKKKKKKKKKKKKKKKKK 44

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :        SPTREMBL_23:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	134	55.1	384	12	Q8JK63	Q8jk63 influenza a
2	134	55.1	566	12	Q98052	Q98052 influenzavi
3	134	55.1	566	12	Q8UZ51	Q8uz51 influenza a
4	134	55.1	566	12	Q8QLN8	Q8qln8 influenza a
5	134	55.1	566	12	Q67132	Q67132 influenza a
6	134	55.1	566	12	Q67125	Q67125 influenzavi
7	134	55.1	566	12	Q8UXR3	Q8uxr3 influenza a
8	134	55.1	566	12	Q91MA7	Q91ma7 influenza a
9	134	55.1	566	12	Q9DHG0	Q9dhg0 influenza a
10	134	55.1	566	12	Q910M5	Q910m5 influenza a
11	134	55.1	566	12	Q67126	Q67126 influenzavi
12	133	54.7	301	12	Q9DXE3	Q9dxe3 influenza a
13	132	54.3	550	12	Q82499	Q82499 influenzavi
14	132	54.3	550	12	Q82498	Q82498 influenzavi
15	132	54.3	550	12	Q82753	Q82753 influenza v
16	132	54.3	566	12	Q82496	Q82496 influenzavi
17	132	54.3	571	12	Q03909	Q03909 influenza a
18	131	53.9	109	12	Q67050	Q67050 influenzavi
19	131	53.9	109	12	Q67053	Q67053 influenzavi
20	131	53.9	109	12	Q67051	Q67051 influenzavi
21	131	53.9	109	12	Q67052	Q67052 influenzavi
22	131	53.9	362	12	Q9QKD3	Q9qkd3 influenza a
23	131	53.9	362	12	Q9QKD1	Q9qkd1 influenza a
24	131	53.9	362	12	Q82513	Q82513 influenzavi
25	131	53.9	362	12	Q9QKD2	Q9qkd2 influenza a
26	131	53.9	362	12	Q84174	Q84174 influenzavi
27	131	53.9	362	12	Q82517	Q82517 influenzavi
28	131	53.9	365	12	Q9DL25	Q9dl25 influenza a
29	131	53.9	367	12	Q9DL22	Q9dl22 influenza a
30	131	53.9	368	12	Q9DL29	Q9dl29 influenza a
31	131	53.9	369	12	Q9DL26	Q9dl26 influenza a
32	131	53.9	369	12	P87689	P87689 influenzavi
33	131	53.9	369	12	Q9DL06	Q9dl06 influenza a
34	131	53.9	371	12	Q9DL24	Q9dl24 influenza a
35	131	53.9	371	12	P87685	P87685 influenzavi
36	131	53.9	373	12	Q9DL20	Q9dl20 influenza a
37	131	53.9	374	12	Q9DL21	Q9dl21 influenza a
38	131	53.9	375	12	Q9DL27	Q9dl27 influenza a
39	131	53.9	375	12	Q9DL05	Q9dl05 influenza a
40	131	53.9	376	12	Q9DL30	Q9dl30 influenza a
41	131	53.9	376	12	Q9DL04	Q9dl04 influenza a
42	131	53.9	377	12	Q9E7P6	Q9e7p6 influenza a
43	131	53.9	382	12	Q9DL03	Q9dl03 influenza a
44	131	53.9	408	12	Q9E7P5	Q9e7p5 influenza a
45	131	53.9	409	12	Q9Q0L5	Q9q0l5 influenza a

ALIGNMENTS

```
RESULT 1
Q8JK63
ID   Q8JK63      PRELIMINARY;      PRT;   384 AA.
AC   Q8JK63;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hemagglutinin (Fragment).
GN   H3HA.
OS   Influenza A virus (A/teal/Germany/wv201r/01).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=205472;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/teal/Germany/wv01r/01;
RA   Werner O., Starick E., Mueller T., Muehle R.;
RT   "Characterisation of avian influenza virus isolates from wild birds
RT   from Germany.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ506781; CAD44999.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER     384    384
SQ   SEQUENCE   384 AA;  42076 MW;  459731795CA5CE38 CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-08;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
```

```
              ||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 2
Q98052
ID   Q98052      PRELIMINARY;      PRT;   566 AA.
AC   Q98052;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin precursor (Fragment).
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81053698; PubMed=6253883;
RA   Both G.W., Sleigh M.J.;
RT   "Complete nucleotide sequence of the haemagglutinin gene from a human
RT   influenza virus of the Hong Kong subtype.";
RL   Nucleic Acids Res. 8:2561-2575(1980).
RN   [2]
RP   SEQUENCE OF 17-344 FROM N.A.
RX   MEDLINE=81194918; PubMed=6164798;
RA   Sleigh M.J., Both G.W., Underwood P.A., Bender V.J.;
RT   "Antigenic drift in the hemagglutinin of the Hong Kong influenza
RT   subtype: Correlation of amino acid changes with alterations in viral
RT   antigenicity.";
RL   J. Virol. 37:845-853(1981).
RN   [3]
RP   SEQUENCE OF 17-566 FROM N.A.
RX   MEDLINE=82033276; PubMed=6169843;
RA   Both G.W., Sleigh M.J.;
RT   "Conservation and variation in the hemagglutinins of Hong Kong subtype
RT   influenza viruses during antigenic drift.";
RL   J. Virol. 39:845-853(1981).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; J02135; AAA43189.1; -.
DR   HSSP; P03437; 1HGE.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    344       POTENTIAL.
FT   CHAIN       346    566       POTENTIAL.
SQ   SEQUENCE   566 AA;  63414 MW;  C447FD465BE4FCF9 CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              ||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 3
Q8UZ51
ID   Q8UZ51      PRELIMINARY;      PRT;   566 AA.
AC   Q8UZ51;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Haemagglutinin.
GN   HA.
OS   Influenza A virus (A/pet bird/Hong Kong/1559/99(H3N8)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=183665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/pet bird/Hong Kong/1559/99;
RA   Chin P., Shortridge K.F.;
RT   "Characterisation of avian H3 influenza viruses.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ427304; CAD20336.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63403 MW;  F11C91B6A0183484 CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              ||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 4
Q8QLN8
ID   Q8QLN8      PRELIMINARY;      PRT;   566 AA.
AC   Q8QLN8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Haemagglutinin.
GN   HA.
OS   Influenza A virus (A/aquatic bird/Hong Kong/399/99(H3N8)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=183664;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/aquatic bird/Hong Kong/399/99;
RA   Chin P., Shortridge K.F.;
RT   "Characterisation of influenza viruses from wild aquatic birds.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ427297; CAD20322.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63412 MW;  68913C222C97B92E CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              ||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368

RESULT 5
```

```
Q67132
ID   Q67132      PRELIMINARY;      PRT;   566 AA.
AC   Q67132;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin.
GN   HA.
OS   Influenza A virus (strain A/Aichi/2/68).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=150147;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Aichi/2/68;
RA   Min J.W., Verhoeyen M., Fang R.-X., Devos R., Huylebroeck D.,
RA   Fiers W.;
RT   "Shift and drift in influenza viruses.";
RL   (In) Carlile M.J., Collins J.F., Moseley B.E. B. (eds.);
RL   SYMPOSIUM OF THE SOCIETY FOR GENERAL MICROBIOLOGY, pp.285-311,
RL   Cambridge University Press, New York (1981).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; M55059; AAA43239.1; -.
DR   HSSP; P03437; 1HGE.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   CHAIN         1    344       HEMAGGLUTININ.
FT   CHAIN       346    566       NEURAMINIDASE.
SQ   SEQUENCE   566 AA;  63441 MW;  E5D1B97DF96FFECA CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 6
Q67125
ID   Q67125      PRELIMINARY;      PRT;   566 AA.
AC   Q67125;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin.
GN   HA.
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Seal/MA/3911/92;
RX   MEDLINE=95146951; PubMed=7844533;
RA   Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT   "The appearance of H3 influenza viruses in seals.";
RL   J. Gen. Virol. 76:199-203(1995).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; L31949; AAA64229.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63456 MW;  AE556302A9EEB99F CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 7
Q8UXR3
ID   Q8UXR3      PRELIMINARY;      PRT;   566 AA.
AC   Q8UXR3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Heamagglutinin.
GN   HA.
OS   Influenza A virus (A/swine/Potsdam/35/82(H3N2)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=183769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/swine/Potsdam/35/82;
RA   Groetzinger I., Suess J., Groetzinger C.;
RT   "Evolution of european human and porcine influenza viruses.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ252132; CAC81018.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63529 MW;  6AA44C84B4DDE68A CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 8
Q91MA7
ID   Q91MA7      PRELIMINARY;      PRT;   566 AA.
AC   Q91MA7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hemagglutinin.
OS   Influenza A virus (A/Hong Kong/1/68(H3N2)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=108859;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Hong Kong/1/68(H3N2);
RX   MEDLINE=21287244; PubMed=11371620;
```

```
RA   Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;
RT   "Pattern of mutation in the genome of influenza A virus on adaptation
RT   to increased virulence in the mouse lung: Identification of functional
RT   themes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AF348176; AAK51718.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63387 MW;  01BB0D465BE158E1 CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
                |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 9
Q9DHG0
ID   Q9DHG0      PRELIMINARY;      PRT;   566 AA.
AC   Q9DHG0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Haemagglutinin precursor.
OS   Influenza A virus H3N2.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=41857;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=clone 7a;
RA   Mohsin M.A., Morris S.J., Smith H., Sweet C.;
RT   "Influenza virus-induced apoptosis: a dual role for viral
RT   neuraminidase.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AJ289703; CAC18525.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin; Signal.
FT   SIGNAL        1     16       POTENTIAL.
SQ   SEQUENCE   566 AA;  63356 MW;  0BA681929300F72F CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
                |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 10
Q910M5
ID   Q910M5      PRELIMINARY;      PRT;   566 AA.
AC   Q910M5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hemagglutinin.
OS   Influenza A virus (A/Hong Kong/1/68(H3N2)).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=108859;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Hong Kong/1/68(H3N2);
RX   MEDLINE=21287244; PubMed=11371620;
RA   Brown E.G., Liu H., Kit L.C., Baird S., Nesrallah M.;
RT   "Pattern of mutation in the genome of influenza A virus on adaptation
RT   to increased virulence in the mouse lung: Identification of functional
RT   themes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:6883-6888(2001).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AF348179; AAK51721.1; -.
DR   EMBL; AF348177; AAK51719.1; -.
DR   EMBL; AF348178; AAK51720.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63530 MW;  7CB9F5BAF1E6E9F4 CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
                |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 11
Q67126
ID   Q67126      PRELIMINARY;      PRT;   566 AA.
AC   Q67126;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin.
GN   HA.
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Seal/MA/3984/92;
RX   MEDLINE=95146951; PubMed=7844533;
RA   Callan R.J., Early G., Kida H., Hinshaw V.S.;
RT   "The appearance of H3 influenza viruses in seals.";
RL   J. Gen. Virol. 76:199-203(1995).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; L32024; AAA64228.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
```

```
KW   Envelope protein; Glycoprotein; Hemagglutinin.
SQ   SEQUENCE   566 AA;  63441 MW;  590576CB4CEE7D08 CRC64;

  Query Match             55.1%;  Score 134;  DB 12;  Length 566;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
                |||||||||||||||||||||||
Db        346 GLFGAIAGFIENGWEGMIDGWYG 368


RESULT 12
Q9DXE3
ID   Q9DXE3     PRELIMINARY;      PRT;   301 AA.
AC   Q9DXE3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinin (Fragment).
GN   HA1.
OS   Influenza A virus (A/Shorebird/Taiwan/31-4/99).
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses; Influenzavirus A.
OX   NCBI_TaxID=140665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Shorebird/Taiwan/31-4/99;
RA   Lee M.S., Cheng P.C., Shien J.H., Cheng M.C., Lee L.H., Shieh H.K.;
RT   "Identification and subtyping of avian influenza virus by reverse
RT   transcription-polymerase chain reaction.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; AF311750; AAG33016.1; -.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER       1      1
FT   NON_TER     301    301
SQ   SEQUENCE   301 AA;  32701 MW;  62A403758B764D57 CRC64;

  Query Match             54.7%;  Score 133;  DB 12;  Length 301;
  Best Local Similarity   95.7%;  Pred. No. 9.7e-08;
  Matches   22;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
                ||||||||||||||||:||||||
Db        250 GLFGAIAGFIENGWEGMVDGWYG 272


RESULT 13
Q82499
ID   Q82499     PRELIMINARY;      PRT;   550 AA.
AC   Q82499;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinins HA1 and HA2 (Fragment).
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82/BS;
RA   Hartley C.A., Ward A.C., Anders E.M.;
RT   "Virulence of influenza virus for mice is associated with loss of
RT   oligosaccharide from the hemagglutinin molecule.";
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; U08859; AAA18782.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER       1      1
FT   CHAIN         1    328       HA1.
FT   CHAIN       330    550       HA2.
SQ   SEQUENCE   550 AA;  61772 MW;  50BD62B6BFE11FD8 CRC64;

  Query Match             54.3%;  Score 132;  DB 12;  Length 550;
  Best Local Similarity   95.7%;  Pred. No. 2.3e-07;
  Matches   22;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
                |:|||||||||||||||||||||
Db        330 GIFGAIAGFIENGWEGMIDGWYG 352


RESULT 14
Q82498
ID   Q82498     PRELIMINARY;      PRT;   550 AA.
AC   Q82498;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hemagglutinins HA1 and HA2 (Fragment).
OS   Influenzavirus A.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   Influenza A viruses.
OX   NCBI_TaxID=197911;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82;
RA   Hartley C.A., Ward A.C., Anders E.M.;
RT   "Virulence of influenza virus for mice is associated with loss of
RT   oligosaccharide from the hemagglutinin molecule.";
RL   Submitted (APR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82;
RX   MEDLINE=88185444; PubMed=3356226;
RA   Nakajima S., Takeuchi Y., Nakajima K.;
RT   "Location on the evolutionary tree of influenza H3 haemagglutinin
RT   genes of Japanese strains isolated during the 1985-6 season.";
RL   Epidemiol. Infect. 100:301-310(1988).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; U08858; AAA18781.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER       1      1
FT   CHAIN         1    328       HA1.
FT   CHAIN       330    550       HA2.
SQ   SEQUENCE   550 AA;  61802 MW;  114413B1CE5A1F6A CRC64;

  Query Match             54.3%;  Score 132;  DB 12;  Length 550;
```

```
  Best Local Similarity   95.7%;  Pred. No. 2.3e-07;
  Matches   22;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |:|||||||||||||||||||||
Db        330 GIFGAIAGFIENGWEGMIDGWYG 352


RESULT 15
Q82753
ID   Q82753      PRELIMINARY;      PRT;   550 AA.
AC   Q82753;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Haemagglutinin (Fragment).
OS   Influenza virus.
OC   Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC   unclassified Orthomyxoviridae.
OX   NCBI_TaxID=11309;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82/BS/ML10;
RX   MEDLINE=97300854; PubMed=9155874;
RA   Hartley C.A., Reading P.C., Ward A.C., Anders E.M.;
RT   "Changes in the hemagglutinin molecule of influenza type A (H3N2)
RT   virus associated with increased virulence for mice.";
RL   Arch. Virol. 142:75-88(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A/Philippines/2/82/BS/ML10;
RX   MEDLINE=97456249; PubMed=9311563;
RA   Ward A.C.;
RT   "Virulence of influenza A virus for mouse lung.";
RL   Virus Genes 14:187-194(1997).
CC   -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC       CELL RECEPTORS AND FOR INITIATING INFECTION (BY SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC       (HA1 AND HA2) LINKED BY A DISULFIDE BOND (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR   EMBL; U08905; AAC79579.1; -.
DR   HSSP; P03437; 2VIU.
DR   InterPro; IPR001364; Hemagglutn.
DR   Pfam; PF00509; Hemagglutinin; 1.
DR   PRINTS; PR00329; HEMAGGLUTN12.
DR   ProDom; PD000225; Hemagglutn; 1.
KW   Envelope protein; Glycoprotein; Hemagglutinin.
FT   NON_TER       1      1
FT   CHAIN         1    328       HAEMAGGLUTININ HA1.
FT   CHAIN       330    550       HAEMAGGLUTININ HA2.
SQ   SEQUENCE   550 AA;  61745 MW;  692A49DE678AC4BC CRC64;

  Query Match             54.3%;  Score 132;  DB 12;  Length 550;
  Best Local Similarity   95.7%;  Pred. No. 2.3e-07;
  Matches   22;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 GLFGAIAGFIENGWEGMIDGWYG 24
              |:|||||||||||||||||||||
Db        330 GIFGAIAGFIENGWEGMIDGWYG 352


Search completed: January 30, 2004, 00:24:41
Job time : 45.446 secs
```

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:06:28 ; Search time 21 Seconds
(without alignments)
36.636 Million cell updates/sec

Title: SEQ10
Perfect score: 38
Sequence: 1 siinfekl 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	100.0	386	1	OACH	ovalbumin [validat
2	34	89.5	627	2	S37954	RNA polymerase I t
3	33	86.8	168	2	F90095	hypothetical prote
4	33	86.8	315	2	T10818	1-aminocyclopropan
5	33	86.8	383	2	S11433	ovalbumin - Japane
6	32	84.2	260	2	G87349	conserved hypothet
7	32	84.2	347	2	A99989	cyclin B [imported
8	32	84.2	1305	2	T23314	hypothetical prote
9	31	81.6	244	2	T06961	ABC transport prot
10	31	81.6	304	2	C64109	site-specific DNA-
11	31	81.6	307	2	E71206	hypothetical prote
12	31	81.6	309	2	H90548	hypothetical prote
13	31	81.6	364	2	A84726	probable poly(ADP-
14	31	81.6	397	2	T34441	hypothetical prote
15	31	81.6	680	2	T42923	infected cell prot
16	30	78.9	142	2	A83093	50S ribosomal prot
17	30	78.9	232	1	DXCH	ovalbumin-related
18	30	78.9	347	2	AB2358	hypothetical prote
19	30	78.9	349	2	H71923	probable inner mem
20	30	78.9	378	2	T18486	hypothetical prote
21	30	78.9	440	2	F70117	hypothetical prote
22	30	78.9	467	2	AF1317	ATP-dependent DNA
23	30	78.9	470	2	D84614	hypothetical prote
24	30	78.9	481	2	B70179	conserved hypothet
25	30	78.9	487	2	C97144	probable membrane
26	30	78.9	520	2	B82206	probable purine-bi
27	30	78.9	558	2	S61604	probable membrane
28	30	78.9	610	2	T25262	hypothetical prote
29	30	78.9	842	2	E72373	hypothetical prote
30	30	78.9	845	2	D90130	hypothetical prote
31	30	78.9	1004	2	B69483	hypothetical prote
32	30	78.9	1115	2	B84476	probable TPR repea
33	30	78.9	1163	2	G97236	ATPase involved in
34	30	78.9	2116	2	T49818	glutamate synthase
35	30	78.9	2469	2	H36812	hypothetical prote
36	30	78.9	4092	1	S38128	dynein heavy chain
37	29	76.3	69	2	A96696	protein F1N21.2 [i
38	29	76.3	88	4	S54022	hypothetical prote
39	29	76.3	110	2	F97094	probable transcrip
40	29	76.3	141	2	H69044	peptidylprolyl iso
41	29	76.3	143	2	S45537	peptidylprolyl iso
42	29	76.3	170	2	G86356	hypothetical prote
43	29	76.3	209	2	D86882	hypothetical prote
44	29	76.3	237	2	F64506	2-hydroxyhepta-2,4
45	29	76.3	275	1	C69030	MJ1225 protein hom

ALIGNMENTS

RESULT 1
OACH
ovalbumin [validated] - chicken
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1979 #sequence_revision 30-Jun-1993 #text_change 15-Sep-2000
C;Accession: A90455; I50402; I50605; A93197; A93827; A90093; A90092; A61297; A42793; A01
R;Woo, S.L.C.; Beattie, W.G.; Catterall, J.F.; Dugaiczyk, A.; Staden, R.; Brownlee, G.G
Biochemistry 20, 6437-6446, 1981
A;Title: Complete nucleotide sequence of the chicken chromosomal ovalbumin gene and its
A;Reference number: A90455; MUID:82069038; PMID:6272839
A;Accession: A90455
A;Molecule type: DNA
A;Residues: 1-386 <WOO>
A;Cross-references: EMBL:V00438; NID:g63719; PIDN:CAA23716.1; PID:g808974
A;Note: a number of silent polymorphic sites are identified and discussed
A;Note: Thr-188 is also predicted
R;Catterall, J.F.; O'Malley, B.W.; Robertson, M.A.; Staden, R.; Tanaka, Y.; Brownlee, G
Nature 275, 510-513, 1978
A;Title: Nucleotide sequence homology at 12 intron-exon junctions in the chick ovalbumin
A;Reference number: I50402; MUID:79010682; PMID:692731
A;Accession: I50402
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-386 <CAT>
A;Cross-references: GB:M34352; NID:g212501; PIDN:AAA48998.1; PID:g212503
R;Robertson, M.A.; Staden, R.; Tanaka, Y.; Catterall, J.F.; O'Malley, B.W.; Brownlee, G.
Nature 278, 370-372, 1979
A;Title: Sequence of three introns in the chick ovalbumin gene.
A;Reference number: I50605; MUID:79135070; PMID:423993
A;Accession: I50605
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4,'A',6-118,'F',120-155 <ROB>
A;Cross-references: EMBL:V00382; NID:g63051; PIDN:CAA23681.1; PID:g63052
R;McReynolds, L.; O'Malley, B.W.; Nisbet, A.D.; Fothergill, J.E.; Givol, D.; Fields, S.,
Nature 273, 723-728, 1978
A;Title: Sequence of chicken ovalbumin mRNA.
A;Reference number: A93197; MUID:78199842; PMID:661981
A;Accession: A93197
A;Molecule type: mRNA
A;Residues: 1-386 <MCR>
A;Cross-references: EMBL:V00383; NID:g63053
A;Note: a minor component has Asp-312
R;Palmiter, R.D.; Gagnon, J.; Walsh, K.A.
Proc. Natl. Acad. Sci. U.S.A. 75, 94-98, 1978
A;Title: Ovalbumin: a secreted protein without a transient hydrophobic leader sequence.
A;Reference number: A93827; MUID:78116057; PMID:272676
A;Accession: A93827
A;Molecule type: protein
A;Residues: 2-33,'X',35-36 <PAL>
R;Thompson, E.O.P.; Fisher, W.K.
Aust. J. Biol. Sci. 31, 443-446, 1978

```
A;Title: A correction and extension of the acetylated amino terminal sequence of ovalbum
A;Reference number: A90093; MUID:79186958; PMID:751625
A;Accession: A90093
A;Molecule type: protein
A;Residues: 2-17 <TH1>
R;Thompson, E.O.P.; Fisher, W.K.
Aust. J. Biol. Sci. 31, 433-442, 1978
A;Title: Amino acid sequences containing half-cystine residues in ovalbumin.
A;Reference number: A90092; MUID:79186957; PMID:751624
A;Accession: A90092
A;Molecule type: protein
A;Residues: 6-17;30-36;61-79;116-124;367-374;380-386 <TH2>
R;Tsunasawa, S.; Narita, K.
J. Biochem. 92, 607-613, 1982
A;Title: Micro-identification of amino-terminal acetylamino acids in proteins.
A;Reference number: A61297; MUID:83056735; PMID:6754709
A;Accession: A61297
A;Molecule type: protein
A;Residues: 2-6 <TSU>
R;Takahashi, N.; Hirose, M.
J. Biol. Chem. 267, 11565-11572, 1992
A;Title: Reversible denaturation of disulfide-reduced ovalbumin and its reoxidation gene
A;Reference number: A42793; MUID:92283876; PMID:1597484
A;Accession: A42793
A;Molecule type: protein
A;Residues: 60-73,'X',75-85;112-119,'EX',122-123 <TAK>
R;Stein, P.E.; Leslie, A.G.W.
submitted to the Brookhaven Protein Data Bank, November 1990
A;Reference number: A50294; PDB:1OVA
A;Contents: annotation; X-ray crystallography, 1.95 angstroms, residues 2-386
R;Stein, P.E.; Leslie, A.G.W.; Finch, J.T.; Carrell, R.W.
J. Mol. Biol. 221, 941-959, 1991
A;Title: Crystal structure of uncleaved ovalbumin at 1.95 Angstroms resolution.
A;Reference number: A58761; MUID:92046044; PMID:1942038
A;Contents: annotation; X-ray crystallography, 1.95 angstroms
C;Genetics:
A;Introns: 56/3; 73/3; 116/3; 156/1; 203/3; 255/3
C;Superfamily: antithrombin III
C;Keywords: acetylated amino end; glycoprotein; phosphoprotein
F;2-384/Product: ovalbumin #status experimental <MAT>
F;2/Modified site: acetylated amino end (Gly) (in mature form) #status experimental
F;69,345/Binding site: phosphate (Ser) (covalent) #status experimental
F;74-121/Disulfide bonds: #status experimental
F;293/Binding site: carbohydrate (Asn) (covalent) #status experimental

  Query Match             100.0%;  Score 38;  DB 1;  Length 386;
  Best Local Similarity   100.0%;  Pred. No. 2.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db        258 SIINFEKL 265


RESULT 2
S37954
RNA polymerase I transcription factor RRN3 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKL125w
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 13-Mar-1998
C;Accession: S37954; S71600
R;Ramezani Rad, M.; Xu, G.; Kirchrath, L.; Fritz, C.; Keuchel, H.; Hollenberg, C.P.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37953
A;Accession: S37954
A;Molecule type: DNA
A;Residues: 1-627 <RAM>
A;Cross-references: EMBL:Z28125; NID:g486210; PID:g486211; MIPS:YKL125w
A;Experimental source: strain S288C
R;Yamamoto, R.T.; Nogi, Y.; Dodd, J.A.; Nomura, M.
EMBO J. 15, 3964-3973, 1996
A;Title: RRN3 gene of Saccharomyces cerevisiae encodes an essential RNA polymerase I tra
A;Reference number: S71600; MUID:96324404; PMID:8670901
A;Accession: S71600
A;Molecule type: DNA
A;Residues: 1-627 <YAM>
C;Genetics:
A;Gene: SGD:RRN3
A;Cross-references: SGD:S0001608; MIPS:YKL125w
A;Map position: 11L
C;Keywords: nucleus

  Query Match             89.5%;  Score 34;  DB 2;  Length 627;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              |||||||
Db        118 SIINFEK 124


RESULT 3
F90095
hypothetical protein orf168 [imported] - Guillardia theta nucleomorph
C;Species: nucleomorph Guillardia theta
A;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: F90095
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei
Nature 410, 1091-1096, 2001
A;Title: The highly reduced genome of an enslaved algal nucleus.
A;Reference number: A99082; MUID:11323671; PMID:11323671
A;Accession: F90095
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-168 <DOU>
A;Cross-references: GB:AF165818; NID:g13794518; PIDN:AAK39893.1; GSPDB:GN00150
C;Genetics:
A;Gene: orf168
A;Map position: 1
A;Genome: nucleomorph
C;Keywords: nucleomorph

  Query Match             86.8%;  Score 33;  DB 2;  Length 168;
  Best Local Similarity   75.0%;  Pred. No. 11;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :||||||:
Db        108 NIINFEKI 115


RESULT 4
T10818
1-aminocyclopropane-1-carboxylate oxidase (EC 1.4.3.-) - kidney bean
C;Species: Phaseolus vulgaris (kidney bean)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 19-May-2000
C;Accession: T10818
R;Pidgeon, C.M.; Facchini, P.J.; Reid, D.M.
submitted to the EMBL Data Library, March 1998
A;Reference number: Z17172
A;Accession: T10818
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-315 <PID>
A;Cross-references: EMBL:AF053354; NID:g3037046; PIDN:AAC12934.1; PID:g3037047
A;Experimental source: cultivar Taylor bush bean; leaf
C;Genetics:
A;Gene: ACO1
C;Superfamily: 1-aminocyclopropane-1-carboxylate oxidase
C;Keywords: ethylene biosynthesis; iron; metalloprotein; oxidoreductase
F;39,177,234/Binding site: iron (His) #status predicted

  Query Match             86.8%;  Score 33;  DB 2;  Length 315;
  Best Local Similarity   85.7%;  Pred. No. 22;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          2 IINFEKL 8
              :||||||
Db          6 VINFEKL 12


RESULT 5
S11433
ovalbumin - Japanese quail
C;Species: Coturnix coturnix japonica (Japanese quail)
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S11433
R;Mucha, J.; Klaudiny, J.; Klaudinyova, V.; Hanes, J.; Simuth, J.
Nucleic Acids Res. 18, 5553, 1990
A;Title: The sequence of Japanese quail ovalbumin cDNA.
A;Reference number: S11433; MUID:91016850; PMID:2216734
A;Accession: S11433
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-383 <MUC>
A;Cross-references: EMBL:X53964; NID:g62643; PIDN:CAA37916.1; PID:g62644
C;Superfamily: antithrombin III

  Query Match             86.8%;  Score 33;  DB 2;  Length 383;
  Best Local Similarity   87.5%;  Pred. No. 27;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              |||:||||
Db        258 SIISFEKL 265


RESULT 6
G87349
conserved hypothetical protein CC0810 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: G87349
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolor
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: G87349
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-260 <STO>
A;Cross-references: GB:AE005673; NID:g13422057; PIDN:AAK22795.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0810

  Query Match             84.2%;  Score 32;  DB 2;  Length 260;
  Best Local Similarity   87.5%;  Pred. No. 29;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              |||||| |
Db        215 SIINFEAL 222


RESULT 7
A99989
cyclin B [imported] - Guillardia theta nucleomorph
C;Species: nucleomorph Guillardia theta
A;Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C;Accession: A99989
R;Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rei
Nature 410, 1091-1096, 2001
A;Title: The highly reduced genome of an enslaved algal nucleus.
A;Reference number: A99082; MUID:11323671; PMID:11323671
A;Accession: A99989
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-347 <DOU>
A;Cross-references: GB:AF165818; NID:g13794469; PIDN:AAK39844.1; GSPDB:GN00150
C;Genetics:
A;Gene: cycB
A;Map position: 1
A;Genome: nucleomorph
C;Keywords: nucleomorph

  Query Match             84.2%;  Score 32;  DB 2;  Length 347;
  Best Local Similarity   62.5%;  Pred. No. 40;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :::|||||
Db         86 NVLNFEKL 93


RESULT 8
T23314
hypothetical protein T14G10.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T23314; T24919
R;Wild, A.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z19725
A;Accession: T23314
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1305 <WIL>
A;Cross-references: EMBL:Z69664; PIDN:CAA93519.1; GSPDB:GN00022; CESP:T14G10.2
A;Experimental source: clone K04D7
R;Wild, A.
submitted to the EMBL Data Library, January 1996
A;Reference number: Z19954
A;Accession: T24919
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1305 <WI2>
A;Cross-references: EMBL:Z68880; PIDN:CAA93100.1; GSPDB:GN00022; CESP:T14G10.2
A;Experimental source: clone T14G10
C;Genetics:
A;Gene: CESP:T14G10.2
A;Map position: 4
A;Introns: 450/1; 463/2; 696/2; 763/2; 843/2; 935/3; 1012/1; 1091/1; 1143/1; 1189/2; 12

  Query Match             84.2%;  Score 32;  DB 2;  Length 1305;
  Best Local Similarity   85.7%;  Pred. No. 1.7e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 IINFEKL 8
              :||||||
Db        903 LINFEKL 909


RESULT 9
T06961
ABC transport protein homolog - Cyanophora paradoxa cyanelle
C;Species: cyanelle Cyanophora paradoxa
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 17-Mar-2000
C;Accession: T06961
R;Stirewalt, V.L.; Michalowski, C.B.; Luffelhardt, W.; Bohnert, H.J.; Bryant, D.A.
submitted to the EMBL Data Library, July 1995
A;Description: Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa.
A;Reference number: Z15840
A;Accession: T06961
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-244 <STI>
A;Cross-references: EMBL:U30821; NID:g1016083; PIDN:AAA81304.1; PID:g1016217
A;Experimental source: strain Pringsheim LB555
```

```
C;Genetics:
A;Genome: cyanelle
C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C;Keywords: cyanelle

  Query Match             81.6%;  Score 31;  DB 2;  Length 244;
  Best Local Similarity   75.0%;  Pred. No. 44;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              |||||::|
Db        204 SIINFDEL 211


RESULT 10
C64109
site-specific DNA-methyltransferase (cytosine-specific) (EC 2.1.1.73) - Haemophilus infl
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Jul-2000
C;Accession: C64109
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: C64109
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-304 <TIGR>
A;Cross-references: GB:U32784; GB:L42023; NID:g3212210; PIDN:AAC22700.1; PID:g1574073; T
C;Superfamily: site-specific methyltransferase (cytosine-specific) EcoRII
C;Keywords: methyltransferase; restriction modification system; S-adenosylmethionine

  Query Match             81.6%;  Score 31;  DB 2;  Length 304;
  Best Local Similarity   85.7%;  Pred. No. 56;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              :||||||
Db        291 AIINFEK 297


RESULT 11
E71206
hypothetical protein PH1919 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: E71206
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: E71206
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-307 <KAW>
A;Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31044.1; PID:g3258361
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1919

  Query Match             81.6%;  Score 31;  DB 2;  Length 307;
  Best Local Similarity   75.0%;  Pred. No. 57;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :|||||:|
Db          3 NIINFEEL 10


RESULT 12
H90548
hypothetical protein MYPU_2960 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: H90548
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: H90548
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-309 <KUR>
A;Cross-references: GB:AL445566; PID:g14089710; PIDN:CAC13469.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_2960
A;Genetic code: SGC3

  Query Match             81.6%;  Score 31;  DB 2;  Length 309;
  Best Local Similarity   71.4%;  Pred. No. 57;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              |::||||
Db        287 SVLNFEK 293


RESULT 13
A84726
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: A84726
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: A84726
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-364 <STO>
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AAD32286.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31860
A;Map position: 2

  Query Match             81.6%;  Score 31;  DB 2;  Length 364;
  Best Local Similarity   85.7%;  Pred. No. 68;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 IINFEKL 8
              ||||:||
Db         88 IINFDKL 94


RESULT 14
T34441
hypothetical protein K11H12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34441
R;Bradshaw, H.
submitted to the EMBL Data Library, February 1997
A;Description: The sequence of C. elegans cosmid K11H12.
A;Reference number: Z21526
A;Accession: T34441
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
```

```
A;Residues: 1-397 <BRA>
A;Cross-references: EMBL:U88168; PIDN:AAC24400.1; GSPDB:GN00022; CESP:K11H12.3
A;Experimental source: strain Bristol N2; clone K11H12
C;Genetics:
A;Gene: CESP:K11H12.3
A;Map position: 4
A;Introns: 39/3; 68/2; 118/1; 206/2; 280/3

  Query Match             81.6%;  Score 31;  DB 2;  Length 397;
  Best Local Similarity   75.0%;  Pred. No. 75;
  Matches    6;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                |||||  |:
Db        190 SIINFSKI 197


RESULT 15
T42923
infected cell protein - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 05-May-2000
C;Accession: T42923
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: Z22274
A;Accession: T42923
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-680 <ALB>
A;Cross-references: EMBL:AF083424; PIDN:AAC95539.1
A;Experimental source: strain 73
C;Genetics:
A;Note: orf07
C;Superfamily: herpesvirus infected cell protein ICP18.5

  Query Match             81.6%;  Score 31;  DB 2;  Length 680;
  Best Local Similarity   71.4%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              |::||||
Db        618 SVLNFEK 624


Search completed: January 30, 2004, 07:09:46
Job time : 23 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: January 30, 2004, 07:06:27 ; Search time 12 Seconds
(without alignments)
31.351 Million cell updates/sec

Title: SEQ10
Perfect score: 38
Sequence: 1 siinfekl 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	100.0	385	1	OVAL_CHICK	P01012 gallus gall
2	34	89.5	627	1	RRN3_YEAST	P36070 saccharomyc
3	33	86.8	382	1	OVAL_COTJA	P19104 coturnix co
4	32	84.2	267	1	RM09_HUMAN	Q9byd2 homo sapien
5	31	81.6	242	1	TPIS_MYCFL	P48779 mycoplasma
6	31	81.6	244	1	YCXD_CYAPA	P48334 cyanophora
7	31	81.6	304	1	MTH5_HAEIN	P45000 haemophilus
8	30	78.9	231	1	RL1_BUCAP	Q8ka67 buchnera ap
9	30	78.9	232	1	OVAX_CHICK	P01013 gallus gall
10	30	78.9	349	1	Y567_HELPJ	Q9zlr4 helicobacte
11	30	78.9	364	1	SERC_DROME	Q9van0 drosophila
12	30	78.9	527	1	PPCK_FUSNN	Q8rei2 fusobacteri
13	30	78.9	558	1	MNT2_YEAST	P53059 saccharomyc
14	30	78.9	842	1	DP3A_THEMA	Q9zhg4 thermotoga
15	30	78.9	1163	1	SBCC_CLOAB	Q97fk1 clostridium
16	30	78.9	2469	1	TEGU_HSVSA	Q01056 herpesvirus
17	30	78.9	4092	1	DYHC_YEAST	P36022 saccharomyc
18	29	76.3	143	1	PPIB_BACSU	P35137 bacillus su
19	29	76.3	237	1	YG56_METJA	Q59050 methanococc
20	29	76.3	394	1	CC91_YEAST	P41733 saccharomyc
21	29	76.3	445	1	RPN5_YEAST	Q12250 saccharomyc
22	29	76.3	453	1	EX7L_RICPR	Q9zcp8 rickettsia
23	29	76.3	482	1	V232_FOWPV	Q9j503 fowlpox vir
24	29	76.3	489	1	T2S3_STAAU	P16667 staphylococ
25	29	76.3	749	1	PCRA_LEUCI	Q9s3q0 leuconostoc
26	29	76.3	1002	1	HPS3_MOUSE	Q91vb4 mus musculu
27	29	76.3	1233	1	SMC1_SCHPO	O94383 schizosacch
28	29	76.3	1234	1	YNX5_CAEEL	P34578 caenorhabdi
29	29	76.3	2054	1	YCF2_PINTH	P41653 pinus thunb
30	28	73.7	102	1	CYT1_ORYSA	P09229 oryza sativ
31	28	73.7	124	1	PA25_AGKHP	O42189 agkistrodon
32	28	73.7	177	1	YA50_METJA	Q58450 methanococc
33	28	73.7	214	1	PYRE_PASMU	Q9cjw4 pasteurella
34	28	73.7	222	1	BID2_YERPE	Q9agd4 yersinia pe
35	28	73.7	240	1	YDED_SCHPO	Q10446 schizosacch
36	28	73.7	242	1	TPIS_MYCHY	P50920 mycoplasma
37	28	73.7	269	1	HIS9_LACLA	Q02150 lactococcus
38	28	73.7	278	1	TNF6_RAT	P36940 rattus norv
39	28	73.7	279	1	TNF6_MOUSE	P41047 mus musculu
40	28	73.7	306	1	COAA_STRP3	Q8k7c7 streptococc
41	28	73.7	306	1	COAA_STRP8	Q8p0v9 streptococc
42	28	73.7	306	1	COAA_STRPY	Q99zh1 streptococc
43	28	73.7	309	1	LDH2_LACPL	P59390 lactobacill
44	28	73.7	315	1	ACC1_LYCES	P05116 lycopersico
45	28	73.7	316	1	ACC2_LYCES	P07920 lycopersico

ALIGNMENTS

RESULT 1
OVAL_CHICK

```
ID   OVAL_CHICK     STANDARD;      PRT;   385 AA.
AC   P01012;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Ovalbumin (Plakalbumin) (Allergen Gal d 2) (Gal d II).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82069038; PubMed=6272839;
RA   Woo S.L.C., Beattie W.G., Catterall J.F., Dugaiczyk A., Staden R.,
RA   Brownlee G.G., O'Malley B.W.;
RT   "Complete nucleotide sequence of the chicken chromosomal ovalbumin
RT   gene and its biological significance.";
RL   Biochemistry 20:6437-6446(1981).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=78199842; PubMed=661981;
RA   McReynolds L., O'Malley B.W., Nisbet A.D., Fothergill J.E., Givol D.,
RA   Fields S., Robertson M., Brownlee G.G.;
RT   "Sequence of chicken ovalbumin mRNA.";
RL   Nature 273:723-728(1978).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=79010682; PubMed=692731;
RA   Catterall J.F., O'Malley B.W., Robertson M.A., Staden R.,
RA   Tanaka Y., Brownlee G.G.;
RT   "Nucleotide sequence homology at 12 intron-exon junctions in the
RT   chick ovalbumin gene.";
RL   Nature 275:510-513(1978).
RN   [4]
RP   SEQUENCE OF 1-35.
RX   MEDLINE=78116057; PubMed=272676;
RA   Palmiter R.D., Gagnon J., Walsh K.A.;
RT   "Ovalbumin: a secreted protein without a transient hydrophobic leader
RT   sequence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 75:94-98(1978).
RN   [5]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=79186958; PubMed=751625;
RA   Thompson E.O.P., Fisher W.K.;
RT   "A correction and extension of the acetylated amino terminal sequence
RT   of ovalbumin.";
RL   Aust. J. Biol. Sci. 31:443-446(1978).
RN   [6]
RP   SEQUENCE OF 5-16; 29-35; 60-78; 115-123; 366-373 AND 379-385.
RX   MEDLINE=79186957; PubMed=751624;
RA   Thompson E.O.P., Fisher W.K.;
RT   "Amino acid sequences containing half-cystine residues in ovalbumin.";
RL   Aust. J. Biol. Sci. 31:433-442(1978).
```

```
RN   [7]
RP   N-TERMINUS ACETYLATION, AND PHOSPHORYLATION OF SER-68; SER-236 AND
RP   SER-240.
RX   MEDLINE=22056091; PubMed=12060738;
RA   MacCoss M.J., McDonald W.H., Saraf A., Sadygov R., Clark J.M.,
RA   Tasto J.J., Gould K.L., Wolters D., Washburn M., Weiss A., Clark J.I.,
RA   Yates J.R. III;
RT   "Shotgun identification of protein modifications from protein
RT   complexes and lens tissue.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:7900-7905(2002).
RN   [8]
RP   X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RX   MEDLINE=90370102; PubMed=2395463;
RA   Stein P.E., Leslie A.G.W., Finch J.T., Turnell W.G., McLaughlin P.J.,
RA   Carrell R.W.;
RT   "Crystal structure of ovalbumin as a model for the reactive centre of
RT   serpins.";
RL   Nature 347:99-102(1990).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS).
RX   MEDLINE=92046044; PubMed=1942038;
RA   Stein P.E., Leslie A.G.W., Finch J.T., Carrell R.W.;
RT   "Crystal structure of uncleaved ovalbumin at 1.95-A resolution.";
RL   J. Mol. Biol. 221:941-959(1991).
RN   [10]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX   MEDLINE=90278960; PubMed=2352279;
RA   Wright H.T., Qian H.X., Huber R.;
RT   "Crystal structure of plakalbumin, a proteolytically nicked form of
RT   ovalbumin. Its relationship to the structure of cleaved alpha-1-
RT   proteinase inhibitor.";
RL   J. Mol. Biol. 213:513-528(1990).
RN   [11]
RP   REVIEW.
RX   MEDLINE=21312433; PubMed=11419711;
RA   Huntington J.A., Stein P.E.;
RT   "Structure and properties of ovalbumin.";
RL   J. Chromatogr. B 756:189-198(2001).
CC   -!- FUNCTION: Not known.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Major protein of egg white.
CC   -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/O/OA.html".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J00895; AAB59956.1; -.
DR   EMBL; V00438; CAA23716.1; -.
DR   EMBL; V00383; CAA23682.1; -.
DR   EMBL; M34352; AAA48998.1; -.
DR   EMBL; M34346; AAA48998.1; JOINED.
DR   EMBL; M34347; AAA48998.1; JOINED.
DR   EMBL; M34348; AAA48998.1; JOINED.
DR   EMBL; M34349; AAA48998.1; JOINED.
DR   EMBL; M34350; AAA48998.1; JOINED.
DR   EMBL; M34351; AAA48998.1; JOINED.
DR   PIR; A90455; OACH.
DR   PDB; 1OVA; 15-JUL-92.
DR   PDB; 1JTI; 18-DEC-02.
DR   PDB; 1VAC; 20-JUN-96.
DR   GlycoSuiteDB; P01012; -.
DR   InterPro; IPR000215; Serpin.
DR   Pfam; PF00079; serpin; 1.
DR   SMART; SM00093; SERPIN; 1.
DR   PROSITE; PS00284; SERPIN; 1.
KW   Serpin; Acetylation; Phosphorylation; Glycoprotein; 3D-structure;
KW   Allergen.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION.
FT   MOD_RES      68     68       PHOSPHORYLATION.
FT   DISULFID     73    120
FT   MOD_RES     236    236       PHOSPHORYLATION.
FT   MOD_RES     240    240       PHOSPHORYLATION.
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .).
FT   MOD_RES     344    344       PHOSPHORYLATION.
FT   ACT_SITE    352    353       REACTIVE BOND HOMOLOG.
FT   VARIANT     311    311       N -> D (IN A MINOR COMPONENT).
FT   CONFLICT    187    187       A -> T (IN REF. 2).
FT   HELIX         3     21
FT   TURN         23     24
FT   STRAND       27     29
FT   HELIX        31     43
FT   TURN         44     44
FT   HELIX        47     57
FT   TURN         58     58
FT   TURN         60     61
FT   TURN         63     64
FT   HELIX        67     70
FT   TURN         71     75
FT   TURN         79     80
FT   HELIX        81     91
FT   STRAND       97    108
FT   TURN        109    110
FT   STRAND      113    113
FT   HELIX       115    124
FT   STRAND      129    132
FT   TURN        135    137
FT   HELIX       138    152
FT   TURN        153    155
FT   TURN        163    164
FT   TURN        168    169
FT   STRAND      172    182
FT   STRAND      184    186
FT   HELIX       190    192
FT   STRAND      194    199
FT   STRAND      205    222
FT   TURN        223    226
FT   STRAND      227    234
FT   TURN        235    236
FT   STRAND      239    246
FT   TURN        249    251
FT   HELIX       252    258
FT   HELIX       261    267
FT   TURN        268    268
FT   TURN        270    272
FT   STRAND      274    283
FT   STRAND      285    292
FT   HELIX       293    299
FT   TURN        300    301
FT   HELIX       304    306
FT   TURN        308    309
FT   TURN        313    315
FT   STRAND      316    316
FT   TURN        319    320
FT   STRAND      325    334
FT   STRAND      338    340
FT   STRAND      357    359
FT   STRAND      364    370
FT   TURN        371    373
FT   STRAND      376    382
SQ   SEQUENCE   385 AA;  42750 MW;  52B4339642388FE3 CRC64;

  Query Match             100.0%;  Score 38;  DB 1;  Length 385;
  Best Local Similarity   100.0%;  Pred. No. 0.88;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db        257 SIINFEKL 264
```

```
RESULT 2
RRN3_YEAST
ID   RRN3_YEAST     STANDARD;      PRT;   627 AA.
AC   P36070;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   RNA polymerase I specific transcription initiation factor RRN3.
GN   RRN3 OR YKL125W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX   MEDLINE=96324404; PubMed=8670901;
RA   Yamamoto R.T., Nogi Y., Dodd J.A., Nomura M.;
RT   "RRN3 gene of Saccharomyces cerevisiae encodes an essential RNA
RT   polymerase I transcription factor which interacts with the polymerase
RT   independently of DNA template.";
RL   EMBO J. 15:3964-3973(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rad M.R., Xu G., Kirchrath L., Fritz C., Keuchel H., Hollenberg C.P.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: REQUIRED FOR EFFICIENT TRANSCRIPTION INITIATION BY RNA
CC       POLYMERASE I. INTERACTS WITH POL I IN THE ABSENCE OF TEMPLATE DNA
CC       AND STIMULATES RECRUITEMENT OF POL I, BUT DOES NOT REMAIN AS PART
CC       OF STABLE PRE-INITIATION COMPLEX.
CC   -!- SUBUNIT: Monomer (Probable).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: TO S.POMBE SPAC18G6.11C AND C.ELEGANS C36E8.1.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z71927; CAA96470.1; -.
DR   EMBL; Z28125; CAA81966.1; -.
DR   PIR; S37954; S37954.
DR   TRANSFAC; T03565; -.
DR   SGD; S0001608; RRN3.
DR   Pfam; PF05327; RRN3; 1.
KW   Transcription regulation; Nuclear protein.
FT   DOMAIN       252    259       POLY-ASP.
FT   DOMAIN       267    274       POLY-ASP.
FT   DOMAIN       277    280       POLY-ASP.
FT   DOMAIN       546    549       POLY-ASN.
SQ   SEQUENCE   627 AA;  72387 MW;  A31E7386A9873FDB CRC64;

  Query Match            89.5%;  Score 34;  DB 1;  Length 627;
  Best Local Similarity  100.0%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              |||||||
Db        118 SIINFEK 124


RESULT 3
OVAL_COTJA
ID   OVAL_COTJA     STANDARD;      PRT;   382 AA.
AC   P19104;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Ovalbumin.
OS   Coturnix coturnix japonica (Japanese quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=93934;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Oviduct;
RX   MEDLINE=91016850; PubMed=2216734;
RA   Mucha J., Klaudiny J., Klaudinyova V., Hanes J., Simuth J.;
RT   "The sequence of Japanese quail ovalbumin cDNA.";
RL   Nucleic Acids Res. 18:5553-5553(1990).
CC   -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X53964; CAA37916.1; -.
DR   PIR; S11433; S11433.
DR   HSSP; P01012; 1OVA.
DR   InterPro; IPR000215; Serpin.
DR   Pfam; PF00079; serpin; 1.
DR   SMART; SM00093; SERPIN; 1.
DR   PROSITE; PS00284; SERPIN; 1.
KW   Serpin; Glycoprotein.
FT   INIT_MET       0      0
FT   DISULFID      73    120       BY SIMILARITY.
FT   CARBOHYD     292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     311    311       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   ACT_SITE     352    353       REACTIVE BOND HOMOLOG.
SQ   SEQUENCE   382 AA;  42108 MW;  A4BB59A1BAE8F316 CRC64;

  Query Match            86.8%;  Score 33;  DB 1;  Length 382;
  Best Local Similarity  87.5%;  Pred. No. 11;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              |||:||||
Db        257 SIISFEKL 264


RESULT 4
RM09_HUMAN
ID   RM09_HUMAN     STANDARD;      PRT;   267 AA.
AC   Q9BYD2; Q9BSW8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   60S ribosomal protein L9, mitochondrial precursor (L9mt).
GN   MRPL9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21293042; PubMed=11279069;
RA   Suzuki T., Terasaki M., Takemoto-Hori C., Hanada T., Ueda T., Wada A.,
RA   Watanabe K.;
RT   "Structural compensation for the deficit of rRNA with proteins in the
RT   mammalian mitochondrial ribosome. Systematic analysis of protein
RT   components of the large ribosomal subunit from mammalian
RT   mitochondria.";
RL   J. Biol. Chem. 276:21724-21736(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932;
```

```
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- SIMILARITY: BELONGS TO THE L9P FAMILY OF RIBOSOMAL PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB049636; BAB40841.1; -.
DR   EMBL; BC004517; AAH04517.1; -.
DR   Genew; HGNC:14277; MRPL9.
DR   InterPro; IPR000244; Ribosomal_L9.
DR   Pfam; PF01281; Ribosomal_L9_N; 1.
KW   Ribosomal protein; Mitochondrion; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION (POTENTIAL).
FT   CHAIN         ?    267       60S RIBOSOMAL PROTEIN L9.
FT   CONFLICT    210    210       A -> E (IN REF. 2).
SQ   SEQUENCE   267 AA;  30185 MW;  346C254220FFD1B4 CRC64;

  Query Match             84.2%;  Score 32;  DB 1;  Length 267;
  Best Local Similarity   71.4%;  Pred. No. 12;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              |::||||
Db        236 SVVNFEK 242


RESULT 5
TPIS_MYCFL
ID   TPIS_MYCFL     STANDARD;      PRT;   242 AA.
AC   P48779;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Triosephosphate isomerase (EC 5.3.1.1) (TIM).
GN   TPIA OR TPI.
OS   Mycoplasma flocculare.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2128;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 27399;
RA   Xiang H., McIntosh M.A.;
RL   Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone
CC       phosphate.
CC   -!- PATHWAY: Plays an important role in several metabolic pathways.
CC   -!- SUBUNIT: Homodimer (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U20509; AAA62167.1; -.
DR   HSSP; P00943; 2BTM.
DR   HAMAP; MF_00147; -; 1.
DR   InterPro; IPR000652; Triophos_ismrse.
DR   Pfam; PF00121; TIM; 1.
DR   ProDom; PD001005; Triophos_ismrse; 1.
DR   TIGRFAMs; TIGR00419; tim; 1.
DR   PROSITE; PS00171; TIM; 1.
KW   Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW   Pentose shunt.
FT   ACT_SITE     98     98       BY SIMILARITY.
FT   ACT_SITE    167    167       BY SIMILARITY.
SQ   SEQUENCE   242 AA;  26969 MW;  B1E8560E2DA18FA1 CRC64;

  Query Match             81.6%;  Score 31;  DB 1;  Length 242;
  Best Local Similarity   71.4%;  Pred. No. 18;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              |::||||
Db        155 SVLNFEK 161


RESULT 6
YCXD_CYAPA
ID   YCXD_CYAPA     STANDARD;      PRT;   244 AA.
AC   P48334;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable ABC transporter ATP-binding protein in ycf23-apcF intergenic
DE   region (ORF244).
OS   Cyanophora paradoxa.
OG   Cyanelle.
OC   Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX   NCBI_TaxID=2762;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UTEX LB 555 / Pringsheim;
RA   Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA   Bryant D.A.;
RT   "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL   Plant Mol. Biol. Rep. 13:327-332(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UTEX LB 555 / Pringsheim;
RA   Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA   Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA   Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT   "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT   the genetic complexity of a primitive plastid.";
RL   (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL   Schwemmler W. (eds.);
RL   Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL   (1997).
CC   -!- SIMILARITY: Belongs to the ABC transporter family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   -----------------------------------------------------------------------
DR   EMBL; U30821; AAA81304.1; -.
DR   PIR; T06961; T06961.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW   Hypothetical protein; ATP-binding; Transport; Cyanelle.
FT   NP_BIND      41     48       ATP (POTENTIAL).
SQ   SEQUENCE   244 AA;  27747 MW;  4C5B357FF9C55D3B CRC64;

  Query Match             81.6%;  Score 31;  DB 1;  Length 244;
  Best Local Similarity   75.0%;  Pred. No. 18;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              |||||::|
Db        204 SIINFDEL 211


RESULT 7
MTH5_HAEIN
ID   MTH5_HAEIN     STANDARD;      PRT;   304 AA.
AC   P45000;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Modification methylase HindV (EC 2.1.1.73) (Cytosine-specific
DE   methyltransferase HindV) (M.HindV).
GN   HINDVM OR HI1041.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC       GRCGYC, CAUSES SPECIFIC METHYLATION ON C-? ON BOTH STRANDS, AND
CC       PROTECTS THE DNA FROM CLEAVAGE BY THE HINDV ENDONUCLEASE.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
CC       adenosyl-L-homocysteine + DNA 5-methylcytosine.
CC   -!- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U32784; AAC22700.1; -.
DR   PIR; C64109; C64109.
DR   HSSP; O14717; 1G55.
DR   REBASE; 3574; M.HindV.
DR   TIGR; HI1041; -.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF00145; DNA_methylase; 1.
DR   PRINTS; PR00105; C5METTRFRASE.
DR   TIGRFAMs; TIGR00675; dcm; 1.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00095; C5_MTASE_2; 1.
KW   Hypothetical protein; Transferase; Methyltransferase;
KW   Restriction system; Complete proteome.
FT   ACT_SITE     75     75       BY SIMILARITY.
SQ   SEQUENCE   304 AA;  34365 MW;  03DA1EAB27C84BBD CRC64;

  Query Match             81.6%;  Score 31;  DB 1;  Length 304;
  Best Local Similarity   85.7%;  Pred. No. 23;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              :||||||
Db        291 AIINFEK 297


RESULT 8
RL1_BUCAP
ID   RL1_BUCAP      STANDARD;      PRT;   231 AA.
AC   Q8KA67;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   50S ribosomal protein L1.
GN   RPLA OR BUSG038.
OS   Buchnera aphidicola (subsp. Schizaphis graminum).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Buchnera.
OX   NCBI_TaxID=98794;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22084549; PubMed=12089438;
RA   Tamas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA   Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT   "50 million years of genomic stasis in endosymbiotic bacteria.";
RL   Science 296:2376-2379(2002).
CC   -!- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND IS
CC       LOCATED IN THE NEIGHBORHOOD OF THE SITE WHERE ELONGATION FACTOR TU
CC       IS BOUND TO THE RIBOSOME (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE L1P FAMILY OF RIBOSOMAL PROTEINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE014080; AAM67609.1; -.
DR   InterPro; IPR005878; L1_bact_chl.
DR   InterPro; IPR002143; Ribosomal_L1.
DR   Pfam; PF00687; Ribosomal_L1; 1.
DR   ProDom; PD001314; Ribosomal_L1; 1.
DR   TIGRFAMs; TIGR01169; rplA_bact; 1.
DR   PROSITE; PS01199; RIBOSOMAL_L1; 1.
KW   Ribosomal protein; rRNA-binding; Complete proteome.
SQ   SEQUENCE   231 AA;  25609 MW;  4D234DACF932A3C2 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 231;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 INFEKL 8
              ||||||
Db         15 INFEKL 20


RESULT 9
```

```
OVAX_CHICK
ID   OVAX_CHICK     STANDARD;      PRT;   232 AA.
AC   P01013;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   Gene X protein (Ovalbumin-related) (Fragment).
GN   X.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81022623; PubMed=7418002;
RA   Heilig R., Perrin F., Gannon F., Mandel J.L., Chambon P.;
RT   "The ovalbumin gene family: structure of the X gene and evolution of
RT   duplicated split genes.";
RL   Cell 20:625-637(1980).
CC   -!- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; J00920; AAA68881.1; -.
DR   EMBL; J00918; AAA68881.1; JOINED.
DR   EMBL; J00919; AAA68881.1; JOINED.
DR   EMBL; V00385; CAA23683.1; -.
DR   EMBL; V00386; CAA23684.1; -.
DR   EMBL; V00387; CAA23685.1; -.
DR   PIR; A01243; DXCH.
DR   HSSP; P01012; 1OVA.
DR   InterPro; IPR000215; Serpin.
DR   Pfam; PF00079; serpin; 1.
DR   SMART; SM00093; SERPIN; 1.
DR   PROSITE; PS00284; SERPIN; 1.
KW   Serpin.
FT   NON_TER       1      1
SQ   SEQUENCE   232 AA;  26291 MW;  6B5B86EC4D3B9195 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 232;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 INFEKL 8
              ||||||
Db        104 INFEKL 109
```

```
RESULT 10
Y567_HELPJ
ID   Y567_HELPJ     STANDARD;      PRT;   349 AA.
AC   Q9ZLR4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein JHP0514.
GN   JHP0514.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Helicobacteraceae; Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE UPF0118 (PERM) FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE001484; AAD06090.1; -.
DR   PIR; H71923; H71923.
DR   InterPro; IPR002549; UPF0118.
DR   Pfam; PF01594; UPF0118; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      6     26       POTENTIAL.
FT   TRANSMEM     27     47       POTENTIAL.
FT   TRANSMEM     56     76       POTENTIAL.
FT   TRANSMEM    143    163       POTENTIAL.
FT   TRANSMEM    195    215       POTENTIAL.
FT   TRANSMEM    224    244       POTENTIAL.
FT   TRANSMEM    258    278       POTENTIAL.
FT   TRANSMEM    300    320       POTENTIAL.
SQ   SEQUENCE   349 AA;  39804 MW;  A1846D48CB3A8B6F CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 349;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 INFEKL 8
              ||||||
Db         85 INFEKL 90
```

```
RESULT 11
SERC_DROME
ID   SERC_DROME     STANDARD;      PRT;   364 AA.
AC   Q9VAN0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable phosphoserine aminotransferase (EC 2.6.1.52) (PSAT).
GN   ESTS:39C10S OR CG11899.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
```

```
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- CATALYTIC ACTIVITY: O-phospho-L-serine + 2-oxoglutarate = 3-
CC       phosphonooxypyruvate + L-glutamate.
CC   -!- COFACTOR: Pyridoxal phosphate.
CC   -!- PATHWAY: REQUIRED BOTH IN MAJOR PHOSPHORYLATED PATHWAY OF SERINE
CC       BIOSYNTHESIS AND IN THE BIOSYNTHESIS OF PYRIDOXINE.
CC   -!- SIMILARITY: Belongs to class-V of pyridoxal-phosphate-dependent
CC       aminotransferases.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE003768; AAF56874.1; -.
DR   HSSP; P23721; 1BJN.
DR   FlyBase; FBgn0014427; ESTS:39C10S.
DR   InterPro; IPR000192; AminotransfV.
DR   InterPro; IPR003248; Pser_amintransf.
DR   Pfam; PF00266; aminotran_5; 1.
DR   ProDom; PD001544; Pser_amintransf; 1.
DR   TIGRFAMs; TIGR01364; serC_1; 1.
DR   PROSITE; PS00595; AA_TRANSFER_CLASS_5; 1.
KW   Serine biosynthesis; Transferase; Aminotransferase;
KW   Pyridoxal phosphate.
FT   BINDING     194    194       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   364 AA;  39540 MW;  DAEA4E2F5BD4DB74 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 364;
  Best Local Similarity   62.5%;  Pred. No. 46;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                ||:|||::
Db        220 SILNFEQM 227


RESULT 12
PPCK_FUSNN
ID   PPCK_FUSNN     STANDARD;      PRT;   527 AA.
AC   Q8REI2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phosphoenolpyruvate carboxykinase [ATP] (EC 4.1.1.49) (PEP
DE   carboxykinase) (Phosphoenolpyruvate carboxylase) (PEPCK).
GN   PCKA OR FN1120.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterales; Fusobacteriaceae;
OC   Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
CC   -!- CATALYTIC ACTIVITY: ATP + oxaloacetate = ADP + phosphoenolpyruvate
CC       + CO(2).
CC   -!- PATHWAY: Rate-limiting gluconeogenic enzyme.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the phosphoenolpyruvate carboxykinase [ATP]
CC       family.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE010616; AAL95316.1; -.
DR   HAMAP; MF_00453; -; 1.
DR   InterPro; IPR001272; PEPCK_ATP.
DR   Pfam; PF01293; PEPCK_ATP; 1.
DR   ProDom; PD004723; PEPCK_ATP; 1.
DR   TIGRFAMs; TIGR00224; pckA; 1.
DR   PROSITE; PS00532; PEPCK_ATP; 1.
KW   Gluconeogenesis; Lyase; Decarboxylase; ATP-binding; Complete proteome.
FT   NP_BIND     230    237       ATP (BY SIMILARITY).
SQ   SEQUENCE   527 AA;  59055 MW;  275849FDF254AC01 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 527;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 IINFEK 7
                ||||||
Db        173 IINFEK 178


RESULT 13
MNT2_YEAST
ID   MNT2_YEAST     STANDARD;      PRT;   558 AA.
AC   P53059;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Alpha-1,3-mannosyltransferase MNT2 (EC 2.4.1.-).
GN   MNT2 OR YGL257C OR NRD558.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=97127827; PubMed=8972578;
RA   Coissac E., Maillier E., Robineau S., Netter P.;
RT   "Sequence of a 39,411 bp DNA fragment covering the left end of
RT   chromosome VII of Saccharomyces cerevisiae.";
RL   Yeast 12:1555-1562(1996).
RN   [2]
```

```
RP   CHARACTERIZATION.
RX   MEDLINE=99453866; PubMed=10521541;
RA   Romero P.A., Lussier M., Veronneau S., Sdicu A.M., Herscovics A.,
RA   Bussey H.;
RT   "Mnt2p and Mnt3p of Saccharomyces cerevisiae are members of the Mnn1p
RT   family of alpha-1,3-mannosyltransferases responsible for adding the
RT   terminal mannose residues of O-linked oligosaccharides.";
RL   Glycobiology 9:1045-1051(1999).
CC   -!- FUNCTION: Mannosyltransferase involved in adding the 4th and 5th
CC       mannose residues of O-linked glycans.
CC   -!- PATHWAY: Glycosylation.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MNN1/MNT FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X94357; CAA64130.1; -.
DR   EMBL; Z72779; CAA96977.1; -.
DR   PIR; S61604; S61604.
DR   SGD; S0003226; MNT2.
DR   GO; GO:0000033; F:alpha-1,3-mannosyltransferase activity; IDA.
DR   GO; GO:0006493; P:O-linked glycosylation; IDA.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW   Signal-anchor; Golgi stack.
FT   DOMAIN        1      6       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      7     27       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       28    558       LUMENAL (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   558 AA;  64852 MW;  3E58ED62B4E291B6 CRC64;

  Query Match              78.9%;  Score 30;  DB 1;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 INFEKL 8
              ||||||
Db        109 INFEKL 114


RESULT 14
DP3A_THEMA
ID   DP3A_THEMA     STANDARD;      PRT;   842 AA.
AC   Q9ZHG4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   DNA polymerase III alpha subunit (EC 2.7.7.7).
GN   DNAE OR TM0461.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99045593; PubMed=9826752;
RA   Huang Y.P., Ito J.;
RT   "The hyperthermophilic bacterium Thermotoga maritima has two different
RT   classes of family C DNA polymerases: evolutionary implications.";
RL   Nucleic Acids Res. 26:5300-5309(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
CC   -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC       RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC       THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC       THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC       epsilon and theta chains) that associates with a tau subunit. This
CC       core dimerizes to form the POLIII' complex. PolIII' associates
CC       with the gamma complex (composed of gamma, delta, delta', psi and
CC       chi chains) and with the beta chain to form the complete DNA
CC       polymerase III complex (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAE
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF063188; AAC80434.1; -.
DR   EMBL; AE001724; AAD35546.1; -.
DR   PIR; E72373; E72373.
DR   TIGR; TM0461; -.
DR   InterPro; IPR003141; PHP_N.
DR   InterPro; IPR004805; PolC_alpha.
DR   SMART; SM00481; POLIIIAc; 1.
DR   TIGRFAMs; TIGR00594; polc; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   Complete proteome.
FT   CONFLICT    377    377       I -> M (IN REF. 1).
SQ   SEQUENCE   842 AA;  96499 MW;  19BB813C8085C277 CRC64;

  Query Match              78.9%;  Score 30;  DB 1;  Length 842;
  Best Local Similarity   62.5%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              |:: ||||
Db         14 SVVRFEKL 21


RESULT 15
SBCC_CLOAB
ID   SBCC_CLOAB     STANDARD;      PRT;  1163 AA.
AC   Q97FK1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Nuclease sbcCD subunit C.
GN   SBCC OR CAC2736.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
```

```
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
CC   -!- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
CC       can inhibit DNA replication and are intermediates in certain DNA
CC       recombination reactions. The complex acts as a 3'->5' double
CC       strand exonuclease that can open hairpins. It also has a 5'
CC       single-strand endonuclease activity (By similarity).
CC   -!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE SMC FAMILY. SBCC SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AE007771; AAK80682.1; -.
DR   PIR; G97236; G97236.
DR   InterPro; IPR003439; ABC_transporter.
KW   Hydrolase; Nuclease; Exonuclease; Endonuclease; DNA replication;
KW   DNA recombination; ATP-binding; Coiled coil; Complete proteome.
FT   NP_BIND      35     42       ATP (POTENTIAL).
FT   DOMAIN      197    415       COILED COIL (POTENTIAL).
FT   DOMAIN      446   1003       COILED COIL (POTENTIAL).
SQ   SEQUENCE   1163 AA;  135507 MW;  CE5F0BD2215D7A92 CRC64;

  Query Match             78.9%;  Score 30;  DB 1;  Length 1163;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          3 INFEKL 8
              ||||||
Db         20 INFEKL 25

Search completed: January 30, 2004, 07:07:49
Job time : 14 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

```
Run on:         January 30, 2004, 07:06:27 ; Search time 70 Seconds
                                             (without alignments)
                                             29.492 Million cell updates/sec

Title:          SEQ10
Perfect score:  38
Sequence:       1 siinfekl 8

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:      830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                1:  sp_archea:*
                2:  sp_bacteria:*
                3:  sp_fungi:*
                4:  sp_human:*
                5:  sp_invertebrate:*
                6:  sp_mammal:*
                7:  sp_mhc:*
                8:  sp_organelle:*
                9:  sp_phage:*
                10:  sp_plant:*
                11:  sp_rodent:*
                12:  sp_virus:*
                13:  sp_vertebrate:*
                14:  sp_unclassified:*
                15:  sp_rvirus:*
                16:  sp_bacteriap:*
                17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	86.8	168	10	Q98RQ3	Q98rq3 guillardia
2	33	86.8	315	10	O65158	O65158 phaseolus v
3	33	86.8	366	2	Q8VQ88	Q8vq88 uncultured
4	32	84.2	260	16	Q9AA01	Q9aa01 caulobacter
5	32	84.2	299	16	Q92KR1	Q92kr1 rhizobium m
6	32	84.2	347	8	Q98RV6	Q98rv6 guillardia
7	32	84.2	443	17	Q8TRP5	Q8trp5 methanosarc
8	32	84.2	444	4	Q8IXU5	Q8ixu5 homo sapien
9	32	84.2	580	4	Q92565	Q92565 homo sapien
10	32	84.2	612	11	Q8BJJ9	Q8bjj9 mus musculu
11	32	84.2	814	11	Q8C0R5	Q8c0r5 mus musculu
12	32	84.2	814	11	Q8C0Q9	Q8c0q9 mus musculu
13	32	84.2	965	2	Q9S4D1	Q9s4d1 staphylococ
14	32	84.2	965	2	Q8VVR0	Q8vvr0 staphylococ
15	32	84.2	1311	5	Q95NL8	Q95nl8 caenorhabdi
16	32	84.2	1347	5	Q95WR8	Q95wr8 caenorhabdi
17	32	84.2	1470	5	Q21218	Q21218 caenorhabdi
18	32	84.2	1573	5	Q9VMF3	Q9vmf3 drosophila
19	32	84.2	1573	5	Q95V18	Q95v18 drosophila
20	31	81.6	97	12	Q8BDR4	Q8bdr4 reindeer pa
21	31	81.6	103	5	Q8IB53	Q8ib53 plasmodium
22	31	81.6	211	5	P91370	P91370 caenorhabdi
23	31	81.6	287	11	Q8R1R1	Q8r1r1 mus musculu
24	31	81.6	307	17	O59582	O59582 pyrococcus
25	31	81.6	309	16	Q98QR5	Q98qr5 mycoplasma
26	31	81.6	364	10	Q9SKB4	Q9skb4 arabidopsis
27	31	81.6	461	11	Q8BZK9	Q8bzk9 mus musculu
28	31	81.6	680	12	Q9YTQ6	Q9ytq6 ateline her
29	31	81.6	834	11	Q8R3E5	Q8r3e5 mus musculu
30	31	81.6	876	11	Q8VCC8	Q8vcc8 mus musculu
31	31	81.6	881	4	Q8WVN0	Q8wvn0 homo sapien
32	31	81.6	881	4	O95634	O95634 homo sapien
33	31	81.6	881	4	O95398	O95398 homo sapien
34	31	81.6	884	11	Q9Z1C8	Q9z1c8 rattus norv
35	31	81.6	1089	5	Q8I605	Q8i605 plasmodium
36	31	81.6	1113	4	Q8TEA3	Q8tea3 homo sapien
37	31	81.6	1138	11	Q8CHG7	Q8chg7 mus musculu
38	31	81.6	1204	4	Q9UHV4	Q9uhv4 homo sapien
39	31	81.6	1391	4	Q8TEU6	Q8teu6 homo sapien
40	31	81.6	1499	4	Q9Y4G8	Q9y4g8 homo sapien
41	31	81.6	1509	4	Q96PC1	Q96pc1 homo sapien
42	31	81.6	1601	4	Q8TEU7	Q8teu7 homo sapien
43	31	81.6	1601	4	Q8NI21	Q8ni21 homo sapien
44	31	81.6	1836	10	Q9LRX4	Q9lrx4 arabidopsis
45	31	81.6	1909	10	Q9LRX3	Q9lrx3 arabidopsis

ALIGNMENTS

```
RESULT 1
Q98RQ3
ID   Q98RQ3      PRELIMINARY;      PRT;   168 AA.
AC   Q98RQ3;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Hypothetical 20.2 kDa protein orf168 from chromosome 1.
GN   ORF168.
OS   Guillardia theta (Cryptomonas phi).
OC   Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX   NCBI_TaxID=55529;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21223349; PubMed=11323671;
RA   Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
RA   Wu X., Reith M., Cavalier-Smith T., Maier U.G.;
RT   "The highly reduced genome of an enslaved algal nucleus.";
RL   Nature 410:1091-1096(2001).
DR   EMBL; AF165818; AAK39893.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   168 AA;  20185 MW;  2874CBD53028A3DD CRC64;

  Query Match             86.8%;  Score 33;  DB 10;  Length 168;
  Best Local Similarity   75.0%;  Pred. No. 40;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :||||||:
Db        108 NIINFEKI 115


RESULT 2
O65158
ID   O65158      PRELIMINARY;      PRT;   315 AA.
AC   O65158;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE   1-aminocyclopropane-1-carboxylic acid oxidase.
GN   ACO1.
OS   Phaseolus vulgaris (Kidney bean) (French bean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX   NCBI_TaxID=3885;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Taylor; TISSUE=Leaf;
RA   Pidgeon C.M., Reid D.M., Facchini P.J.;
RT   "Light induced changes in ethylene production in Phaseolus vulgaris
RT   cv. Taylor.";
RL   Plant Physiol. 114:165-165(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Taylor; TISSUE=Leaf;
RA   Pidgeon C.M., Facchini P.J., Reid D.M.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF053354; AAC12934.1; -.
DR   InterPro; IPR005123; 2OG-FeII_Oxy.
DR   Pfam; PF03171; 2OG-FeII_Oxy; 1.
SQ   SEQUENCE   315 AA;  36141 MW;  6898811DF88E9BB1 CRC64;

  Query Match             86.8%;  Score 33;  DB 10;  Length 315;
  Best Local Similarity   85.7%;  Pred. No. 72;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 IINFEKL 8
               :||||||
Db          6 VINFEKL 12


RESULT 3
Q8VQ88
ID   Q8VQ88      PRELIMINARY;      PRT;   366 AA.
AC   Q8VQ88;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Formyltetrahydrofolate synthetase (Fragment).
OS   uncultured environmental landfill bacterium.
OC   Bacteria; environmental samples.
OX   NCBI_TaxID=184235;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Uz I., Lindner A.S., Rasche M.E., Townsend T.G., Ogram A.V.;
RT   "Microbial Characterization of Landfill Cover Soils of Different
RT   Ages.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF459462; AAL67505.1; -.
DR   InterPro; IPR000559; Fmtethyd_synth.
DR   Pfam; PF01268; FTHFS; 1.
FT   NON_TER       1      1
FT   NON_TER     366    366
SQ   SEQUENCE   366 AA;  39816 MW;  EB4685D501E97026 CRC64;

  Query Match             86.8%;  Score 33;  DB 2;  Length 366;
  Best Local Similarity   85.7%;  Pred. No. 83;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 IINFEKL 8
              |:|||||
Db        346 IVNFEKL 352


RESULT 4
Q9AA01
ID   Q9AA01      PRELIMINARY;      PRT;   260 AA.
AC   Q9AA01;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein CC0810.
GN   CC0810.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC   Caulobacteraceae; Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005757; AAK22795.1; -.
DR   TIGR; CC0810; -.
DR   InterPro; IPR005269; Cons_hypoth730.
DR   Pfam; PF03641; Lysine_decarbox; 1.
DR   TIGRFAMs; TIGR00730; TIGR00730; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   260 AA;  28789 MW;  E4CF85223B12EF1B CRC64;

  Query Match             84.2%;  Score 32;  DB 16;  Length 260;
  Best Local Similarity   87.5%;  Pred. No. 98;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                |||||| |
Db        215 SIINFEAL 222


RESULT 5
Q92KR1
ID   Q92KR1      PRELIMINARY;      PRT;   299 AA.
AC   Q92KR1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein R00434.
GN   R00434 OR SMC01731.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR   EMBL; AL591783; CAC41871.1; -.
DR   InterPro; IPR005269; Cons_hypoth730.
DR   Pfam; PF03641; Lysine_decarbox; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   299 AA;  33440 MW;  9BD61375BE91F195 CRC64;

  Query Match             84.2%;  Score 32;  DB 16;  Length 299;
  Best Local Similarity   87.5%;  Pred. No. 1.1e+02;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                |||||| |
```

```
Db        246 SIINFEAL 253


RESULT 6
Q98RV6
ID   Q98RV6      PRELIMINARY;      PRT;   347 AA.
AC   Q98RV6;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Cyclin B.
GN   CYCB.
OS   Guillardia theta (Cryptomonas phi).
OG   Nucleomorph.
OC   Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX   NCBI_TaxID=55529;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21223349; PubMed=11323671;
RA   Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
RA   Wu X., Reith M., Cavalier-Smith T., Maier U.G.;
RT   "The highly reduced genome of an enslaved algal nucleus.";
RL   Nature 410:1091-1096(2001).
CC   -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY.
DR   EMBL; AF165818; AAK39844.1; -.
DR   InterPro; IPR006670; Cyclin.
DR   InterPro; IPR006671; Cyclin_N.
DR   Pfam; PF00134; cyclin; 1.
DR   SMART; SM00385; CYCLIN; 2.
DR   PROSITE; PS00292; CYCLINS; 1.
KW   Cell cycle; Cell division; Cyclin.
SQ   SEQUENCE   347 AA;  41616 MW;  CC579909C403A04D CRC64;

  Query Match             84.2%;  Score 32;  DB 8;  Length 347;
  Best Local Similarity   62.5%;  Pred. No. 1.3e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :::|||||
Db         86 NVLNFEKL 93


RESULT 7
Q8TRP5
ID   Q8TRP5      PRELIMINARY;      PRT;   443 AA.
AC   Q8TRP5;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Iron ABC transporter, solute-binding protein.
GN   MA1130.
OS   Methanosarcina acetivorans.
OC   Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2214;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C2A / ATCC 35395 / DSM 2834;
RX   MEDLINE=21929760; PubMed=11932238;
RA   Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA   FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA   Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA   Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA   Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA   Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA   Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA   Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA   Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA   Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA   Metcalf W.W., Birren B.;
RT   "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT   and physiological diversity.";
RL   Genome Res. 12:532-542(2002).
DR   EMBL; AE010779; AAM04551.1; -.
DR   InterPro; IPR002491; Peripla_BP.
DR   Pfam; PF01497; Peripla_BP_2; 1.
KW   Complete proteome.
SQ   SEQUENCE   443 AA;  49537 MW;  0D3F2198B20E88C9 CRC64;

  Query Match             84.2%;  Score 32;  DB 17;  Length 443;
  Best Local Similarity   75.0%;  Pred. No. 1.6e+02;
  Matches    6;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :|||:|||
Db        166 TIINYEKL 173


RESULT 8
Q8IXU5
ID   Q8IXU5      PRELIMINARY;      PRT;   444 AA.
AC   Q8IXU5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to guanine nucleotide exchange factor for Rap1,
DE   M-Ras-regulated GEF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC039203; AAH39203.1; -.
SQ   SEQUENCE   444 AA;  52051 MW;  ECBEDF820DDFF13B CRC64;

  Query Match             84.2%;  Score 32;  DB 4;  Length 444;
  Best Local Similarity   62.5%;  Pred. No. 1.6e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :::|||||
Db        381 NLVNFEKL 388


RESULT 9
Q92565
ID   Q92565      PRELIMINARY;      PRT;   580 AA.
AC   Q92565;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein KIAA0277.
GN   KIAA0277.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,
RA   Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
DR   EMBL; D87467; BAA13406.1; -.
DR   InterPro; IPR000651; RasGEFN.
DR   InterPro; IPR001895; RasGRF_CDC25.
DR   Pfam; PF00617; RasGEF; 1.
DR   Pfam; PF00618; RasGEFN; 1.
```

```
DR   SMART; SM00147; RasGEF; 1.
DR   SMART; SM00229; RasGEFN; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   580 AA;  67733 MW;  732FB7AA11DFDA1C CRC64;

  Query Match             84.2%;  Score 32;  DB 4;  Length 580;
  Best Local Similarity   62.5%;  Pred. No. 2.1e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :::|||||
Db        517 NLVNFEKL 524


RESULT 10
Q8BJJ9
ID   Q8BJJ9      PRELIMINARY;      PRT;   612 AA.
AC   Q8BJJ9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical guanine-nucleotide dissociation stimulators CDC25
DE   family/guanine nucleotide exchange factor for Ras-like GTPases;
DE   N-terminal motif containing protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK083591; BAC38963.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   612 AA;  70913 MW;  DBD6552DDFD392B6 CRC64;

  Query Match             84.2%;  Score 32;  DB 11;  Length 612;
  Best Local Similarity   62.5%;  Pred. No. 2.2e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :::|||||
Db        548 NLVNFEKL 555


RESULT 11
Q8C0R5
ID   Q8C0R5      PRELIMINARY;      PRT;   814 AA.
AC   Q8C0R5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical guanine-nucleotide dissociation stimulators CDC25
DE   family/guanine nucleotide exchange factor for Ras-like GTPases;
DE   N-terminal motif containing protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK029995; BAC26723.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   814 AA;  93681 MW;  A64900A72AA3E89B CRC64;

  Query Match             84.2%;  Score 32;  DB 11;  Length 814;
  Best Local Similarity   62.5%;  Pred. No. 2.9e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :::|||||
Db        750 NLVNFEKL 757


RESULT 12
Q8C0Q9
ID   Q8C0Q9      PRELIMINARY;      PRT;   814 AA.
AC   Q8C0Q9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical guanine-nucleotide dissociation stimulators CDC25
DE   family/guanine nucleotide exchange factor for Ras-like GTPases;
DE   N-terminal motif containing protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK030016; BAC26736.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   814 AA;  93754 MW;  AA04EEC74CE13735 CRC64;

  Query Match             84.2%;  Score 32;  DB 11;  Length 814;
  Best Local Similarity   62.5%;  Pred. No. 2.9e+02;
  Matches    5;  Conservative    3;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              :::|||||
Db        750 NLVNFEKL 757


RESULT 13
Q9S4D1
ID   Q9S4D1      PRELIMINARY;      PRT;   965 AA.
AC   Q9S4D1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Lantibiotic modifying enzyme.
OS   Staphylococcus aureus.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C55;
RX   MEDLINE=99346225; PubMed=10417203;
RA   Navaratna M.A., Sahl H.G., Tagg J.R.;
RT   "Identification of genes encoding two-component lantibiotic production
RT   in Staphylococcus aureus C55 and other phage group II S. aureus
RT   strains and demonstration of an association with the exfoliative toxin
RT   B gene.";
RL   Infect. Immun. 67:4268-4271(1999).
DR   EMBL; AF147744; AAD47013.1; -.
SQ   SEQUENCE   965 AA;  111570 MW;  BC1E4DA8DAC4F346 CRC64;
```

```
  Query Match             84.2%;  Score 32;  DB 2;  Length 965;
  Best Local Similarity   71.4%;  Pred. No. 3.4e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              |::||||
Db        202 SVVNFEK 208


RESULT 14
Q8VVR0
ID   Q8VVR0      PRELIMINARY;      PRT;   965 AA.
AC   Q8VVR0;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   ORF42.
OS   Staphylococcus aureus.
OG   Plasmid ETB plasmid.
OC   Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX   NCBI_TaxID=1280;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TY4;
RX   MEDLINE=21562640; PubMed=11705958;
RA   Yamaguchi T., Hayashi T., Takami H., Ohnisi M., Murata T.,
RA   Nakayama K., Asakawa K., Ohara M., Komatsuzawa H., Sugai M.;
RT   "Complete Nucleotide Sequence of a Staphylococcus aureus Exfoliative
RT   Toxin B Plasmid and Identification of a Novel ADP-Ribosyltransferase,
RT   EDIN-C.";
RL   Infect. Immun. 69:7760-7771(2001).
DR   EMBL; AP003088; BAB78440.1; -.
KW   Plasmid.
SQ   SEQUENCE   965 AA;  111561 MW;  C08E9FD8DAC5F82B CRC64;

  Query Match             84.2%;  Score 32;  DB 2;  Length 965;
  Best Local Similarity   71.4%;  Pred. No. 3.4e+02;
  Matches    5;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEK 7
              |::||||
Db        202 SVVNFEK 208


RESULT 15
Q95NL8
ID   Q95NL8      PRELIMINARY;      PRT;  1311 AA.
AC   Q95NL8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   T14G10.2b protein (PXF isoform B).
GN   T14G10.2 OR T14G10.2B.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wild A.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Verheijen M., van Berkel W., Jansen G., de Rooij J., Plasterk R.H.,
RA   Bos J.L., Zwartkruis F.J.T.;
RT   "Characterization of pxf, the C. elegans homolog of human PDZ-GEFs.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Wild A.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; Z68880; CAC42342.1; -.
DR   EMBL; Z69664; CAC42342.1; JOINED.
DR   EMBL; AF308448; AAL09434.1; -.
DR   EMBL; Z69664; CAC42313.1; -.
DR   EMBL; Z68880; CAC42313.1; JOINED.
DR   WormPep; T14G10.2b; CE28081.
DR   InterPro; IPR000595; cNMP_binding.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR000651; RasGEFN.
DR   InterPro; IPR001895; RasGRF_CDC25.
DR   InterPro; IPR000159; RA_domain.
DR   Pfam; PF00027; cNMP_binding; 2.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00788; RA; 1.
DR   Pfam; PF00617; RasGEF; 1.
DR   SMART; SM00100; cNMP; 1.
DR   SMART; SM00228; PDZ; 1.
DR   SMART; SM00314; RA; 1.
DR   SMART; SM00147; RasGEF; 1.
DR   SMART; SM00229; RasGEFN; 1.
DR   PROSITE; PS50042; CNMP_BINDING_3; 1.
DR   PROSITE; PS50106; PDZ; 1.
SQ   SEQUENCE   1311 AA;  147003 MW;  A4870F07DC201F97 CRC64;

  Query Match             84.2%;  Score 32;  DB 5;  Length 1311;
  Best Local Similarity   85.7%;  Pred. No. 4.5e+02;
  Matches    6;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          2 IINFEKL 8
              :||||||
Db       1068 LINFEKL 1074


Search completed: January 30, 2004, 07:09:11
Job time : 71 secs
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         January 30, 2004, 07:15:53 ; Search time 71 Seconds
                                             (without alignments)
                                             17.885 Million cell updates/sec

Title:          SEQ10
Perfect score:  38
Sequence:       1 siinfekl 8

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters:      125

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
                 Maximum Match 100%
                 Listing first 250 summaries

Database :       A_Geneseq_19Jun03:*
                1:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                2:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                3:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                4:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                5:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                6:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                7:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                8:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                9:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                10:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                11:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                12:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                13:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                14:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                15:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                16:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                17:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                18:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                19:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                20:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                21:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                22:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                23:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
                24:  /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38	100.0	8	15	AAR57996	Ova257-264. Synth
2	38	100.0	8	16	AAR83938	MHC class I restri
3	38	100.0	8	17	AAR89157	Peptide Ova8 used
4	38	100.0	8	18	AAW19955	Chicken OVA-peptid
5	38	100.0	8	18	AAW14087	MHC class I molecu
6	38	100.0	8	18	AAW04642	Ovalbumin-derived
7	38	100.0	8	19	AAW80296	Amino acids 257-26
8	38	100.0	8	19	AAW70375	Ovalbumin peptide
9	38	100.0	8	19	AAW68308	MHC binding peptid
10	38	100.0	8	19	AAW68365	MHC binding peptid
11	38	100.0	8	19	AAW60700	Ovalbumin peptide
12	38	100.0	8	19	AAW54265	Ovalbumin peptide
13	38	100.0	8	19	AAW52838	Ovalbumin artifici
14	38	100.0	8	20	AAY42307	Ovalbumin-derived
15	38	100.0	8	20	AAY16871	OVA peptide sequen
16	38	100.0	8	20	AAY03780	Ovalbumin peptide
17	38	100.0	8	20	AAW99480	Ovalbumin-derived
18	38	100.0	8	20	AAW67585	T-cell activation
19	38	100.0	8	21	AAB13763	Peptide fragment f
20	38	100.0	8	21	AAB29465	Negative control p
21	38	100.0	8	21	AAB26484	Tumour associated
22	38	100.0	8	21	AAB13785	OVA-derived peptid
23	38	100.0	8	21	AAY68311	Altered MHC determ
24	38	100.0	8	21	AAY59401	Ovalbumin protein
25	38	100.0	8	21	AAY52564	Murine ovalbumin M
26	38	100.0	8	21	AAY52965	Altered MHC determ
27	38	100.0	8	22	AAG77871	Ovalbumin 257-264
28	38	100.0	8	22	AAM52562	Cytotoxic T-cell e
29	38	100.0	8	22	AAE13119	Ovalbumin (OVA)-de
30	38	100.0	8	22	AAE12145	Murine ovalbumin (
31	38	100.0	8	22	AAE09514	Human ovalbumin pe
32	38	100.0	8	22	AAG63855	Amino acid sequenc
33	38	100.0	8	22	AAG66422	Chicken ovalbumin
34	38	100.0	8	22	AAU05398	Peptide released f
35	38	100.0	8	22	AAE06033	Chicken ovalbumin
36	38	100.0	8	22	AAB84316	Peptide used to pr
37	38	100.0	8	22	AAB99354	Ovalbumin cytotoxi
38	38	100.0	8	22	AAB82176	Immunodominant CTL
39	38	100.0	8	22	AAB81122	Chicken ovalbumin
40	38	100.0	8	22	AAB82065	Ovalbumin-derived
41	38	100.0	8	22	AAB92374	Miscellaneous pept
42	38	100.0	8	22	AAB48950	Ovalbumin MHC clas
43	38	100.0	8	23	AAE28959	Chicken ovalbumin
44	38	100.0	8	23	AAE26368	Ovalbumin CTL epit
45	38	100.0	8	23	ABB79933	Ovalbumin T-cell e
46	38	100.0	8	23	ABG93028	Mouse class 1 MHC
47	38	100.0	8	23	AAE25400	Ovalbumin peptide
48	38	100.0	8	23	ABG31661	Chicken ovalbumin
49	38	100.0	8	23	ABG31967	Chicken ovalbumin
50	38	100.0	8	23	AAU99718	Mouse MHC classI K
51	38	100.0	8	23	ABB08108	Chicken ovalbumin
52	38	100.0	8	23	ABB81273	Chicken OVA 257-26
53	38	100.0	8	23	AAE22531	Ovalbumin peptide
54	38	100.0	8	23	ABB76050	Ovalbumin, H-2kb r
55	38	100.0	8	23	AAE19945	Cytotoxic T-cell e
56	38	100.0	8	23	ABB09907	Ovalbumin peptide
57	38	100.0	8	23	AAU76942	OVA peptide (257-2
58	38	100.0	8	23	AAU76802	MHC class I-restri
59	38	100.0	8	23	AAU76869	OVA peptide fragme
60	38	100.0	8	23	AAU11866	Ovalbumin drived c
61	38	100.0	8	23	AAU11239	Immunodominant Kb-
62	38	100.0	8	23	AAU09820	Ovalbumin-derived
63	38	100.0	8	23	AAE13215	Hen egg ovalbumin
64	38	100.0	8	23	AAE13436	Chicken ovalbumin
65	38	100.0	8	23	AAU75056	Ovalbumin antigeni
66	38	100.0	8	24	ABU07743	Chicken ovalbumin
67	38	100.0	8	24	ABU08619	Ovalbumin (OVA) re
68	38	100.0	8	24	ABP57401	Synthetic 8mer pep
69	38	100.0	8	24	ABP58359	Ovalbumin-derived
70	38	100.0	8	24	ABG73081	MHC Class I peptid
71	38	100.0	8	24	ABP56760	Ovalbumin derived
72	38	100.0	8	24	ABP60027	Ovalbumin antigeni
73	38	100.0	8	24	ABU11029	Ovalbumin immunodo
74	38	100.0	9	22	AAB84323	Peptide used to pr
75	38	100.0	9	24	ABP57402	Synthetic 9mer pep
76	38	100.0	10	18	AAW04643	Ovalbumin-derived
77	38	100.0	10	18	AAW04644	Ovalbumin-derived
78	38	100.0	10	23	AAU09821	Modified ovalbumin
79	38	100.0	10	23	AAU09825	Modified ovalbumin
80	38	100.0	12	18	AAW14122	OVA protein derive
81	38	100.0	12	23	AAU09822	Modified ovalbumin
82	38	100.0	12	23	AAU09826	Modified ovalbumin

83	38	100.0	12	23	AAU09827	Modified ovalbumin
84	38	100.0	14	23	ABB76049	Peptide insert in
85	38	100.0	14	23	AAU09823	Modified ovalbumin
86	38	100.0	14	23	AAU09828	Modified ovalbumin
87	38	100.0	15	23	AAU09824	Modified ovalbumin
88	38	100.0	16	24	ABP57403	Synthetic 16mer pe
89	38	100.0	19	18	AAW19957	BiP-binding domain
90	38	100.0	19	18	AAW19956	OVA-BiP-binding do
91	38	100.0	19	23	AAE13446	Chicken MHC class
92	38	100.0	19	23	AAE13447	Chicken MHC class
93	38	100.0	19	24	ABP57404	Synthetic 19mer pe
94	38	100.0	24	14	AAR32294	Sequence of synthe
95	38	100.0	24	14	AAR41450	Antigenic peptide
96	38	100.0	24	18	AAW04645	Ovalbumin-derived
97	38	100.0	24	22	AAG65170	Ovalbumin based pe
98	38	100.0	24	22	AAB74439	Myelin basic prote
99	38	100.0	24	23	ABG31664	Ovalbumin (OVA) pe
100	38	100.0	26	24	ABP57405	Synthetic 26mer pe
101	38	100.0	26	24	ABP57406	Synthetic 26mer* p
102	38	100.0	30	23	AAE13448	Chicken MHC class
103	38	100.0	31	24	ABP57407	Synthetic 31mer pe
104	38	100.0	35	18	AAW04646	Ovalbumin-derived
105	38	100.0	36	24	AAO26741	Chicken ovalbumin
106	38	100.0	43	22	AAB84325	Amino acid sequenc
107	38	100.0	47	22	AAB84321	Amino acid sequenc
108	38	100.0	48	22	AAB84322	Amino acid sequenc
109	38	100.0	49	22	AAB48953	Tn5-DICE ovalbumin
110	38	100.0	57	22	AAB48954	DICE-I ovalbumin M
111	38	100.0	100	23	AAE13458	Chicken ovalbumin
112	38	100.0	100	23	AAE13460	Chicken ovalbumin
113	38	100.0	103	23	AAE13459	Chicken ovalbumin
114	38	100.0	103	23	AAE13461	Chicken ovalbumin
115	38	100.0	106	17	AAR89966	Polytope sequence.
116	38	100.0	108	23	AAE13462	Chicken ovalbumin
117	38	100.0	111	23	AAE13463	Chicken ovalbumin
118	38	100.0	132	21	AAY52575	Amino acid sequenc
119	38	100.0	386	23	AAE13435	Chicken ovalbumin
120	38	100.0	409	22	AAB31545	Amino acid sequenc
121	38	100.0	479	22	AAE13112	Human HER300-rGM-C
122	38	100.0	541	23	AAU99725	Yeast/mouse SS-OVA
123	38	100.0	564	22	AAE13110	Human HER500 fusio
124	38	100.0	697	22	AAE13111	Human HER500-rGM-C
125	38	100.0	948	22	AAB31611	Amino acid sequenc

ALIGNMENTS

```
RESULT 1
AAR57996
ID   AAR57996 standard; Protein; 8 AA.
XX
AC   AAR57996;
XX
DT   25-MAR-2003  (updated)
DT   30-MAR-1995  (first entry)
XX
DE   Ova257-264.
XX
KW   Ova; ovalbumin; cytosol; cytolytic immune response;
KW   vaccinia virus; promoter.
XX
OS   Synthetic.
XX
PN   WO9417816-A1.
XX
PD   18-AUG-1994.
XX
PF   27-JAN-1994;    94WO-US01183.
XX
PR   10-FEB-1993;    93US-0016066.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Goldberg AL,  Rock KL;
XX
DR   WPI; 1994-279383/34.
XX
PT   Method for blocking cytolytic immune responses - useful for
PT   treatment of autoimmune diseases and preventing organ and graft
PT   rejection
XX
PS   Example 2; Page 44; 89pp; English.
XX
CC   Ova257-264 was constructed by inserting a synthetic oligonucleotide
CC   (AAQ67342) behind the vaccinia virus p7.5 early/late promoter in pSc11
CC   which was modified such that the restriction sites SalI and NotI
CC   were substituted for the SmaI site.  The oligonucleotide consists of
CC   a SalI site, Kozak's consensus sequence for efficient translation,
CC   an initiation codon, nucleotides encoding the peptide given in
CC   AAR57996, two stop codons, and a NotI site.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   8 AA;

  Query Match           100.0%;  Score 38;  DB 15;  Length 8;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 2
AAR83938
ID   AAR83938 standard; peptide; 8 AA.
XX
AC   AAR83938;
XX
DT   05-JUN-1996  (first entry)
XX
DE   MHC class I restricted antigenic peptide #8.
XX
KW   MHC class I; antigen; MAGE; melanoma; breast cancer; bladder cancer;
KW   Titermax; cytotoxic T-lymphocyte; tumour; pathogenic disease; bacteria;
KW   parasite; human; animal.
XX
OS   Synthetic.
XX
PN   WO9528958-A1.
XX
PD   02-NOV-1995.
XX
PF   21-APR-1995;    95WO-US04975.
XX
PR   22-APR-1994;    94US-0233496.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Dyall R,  Nikolic-Zugic J;
XX
DR   WPI; 1995-382848/49.
XX
PT   Cytotoxic T-cell induction by MHC class I-restricted peptide in
PT   adjuvant - useful for treating tumours and bacterial or parasitic
PT   pathogenic diseases
XX
PS   Claim 11; Page 38; 50pp; English.
XX
CC   The sequences given in AAR83931-49 are MHC class I restricted 8-12
CC   amino acid antigenic peptides.  This peptide represents ovalbumin
CC   residues 257-264.  These peptides may be administered to a subject
CC   in combination with a suitable adjuvant, pref. Titermax (RTM), to
CC   induce cytotoxic T-lymphocytes.  This method may be used in the
```

```
CC   treatment of a tumour or a pathogenic disease, esp. diseases of
CC   bacterial or parasitic origin, in humans and animals, e.g monkeys,
CC   dogs, cows, horses, etc.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 16;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 3
AAR89157
ID   AAR89157 standard; peptide; 8 AA.
XX
AC   AAR89157;
XX
DT   25-MAR-2003  (updated)
DT   03-SEP-1996  (first entry)
XX
DE   Peptide Ova8 used in an MHC stripping/reloading method.
XX
KW   Antigen; major histocompatibility complex; cell surface; stimulation;
KW   cytotoxic T lymphocyte; CTL; endogenous; exogenous; peptide; spleen;
KW   allelic restriction; peripheral blood lymphocytes; ganglion; placenta;
KW   native form; infection; tumour; autoimmune disease.
XX
OS   Synthetic.
XX
PN   WO9601891-A1.
XX
PD   25-JAN-1996.
XX
PF   06-JUL-1995;   95WO-FR00907.
XX
PR   07-JUL-1994;   94FR-0008427.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
XX
PI   Langlade Demoyen P,  Kourilsky P,  Abastado J;
XX
DR   WPI; 1996-097621/10.
XX
PT   Cell population with high surface density of exogenous peptide bound
PT   to MHC molecules  - prepd. by stripping endogenous peptide and
PT   reloading exogenous peptide(s), useful for stimulating cytotoxic
PT   lymphocytes in cases of infection, tumour and auto-immune disease
XX
PS   Disclosure; Page 9; 37pp; French.
XX
CC   Peptides AAR89154-73 are examples of exogenous peptides which are
CC   "loaded" onto the antigen-presenting major histocompatibility complex
CC   (MHC) on the cell surface in a novel method of stimulating cytotoxic T
CC   lymphocytes. The method involves treating cells, e.g. at a pH 5 or below
CC   or pH 9 or above, to remove the endogenous peptides from the MHC,
CC   followed by recharging the complexes with specified exogenous peptides
CC   having the same allelic restriction as the MHC. The cells are pref.
CC   peripheral blood lymphocytes, spleen, ganglion or placental cells which
CC   are able to present the exogenous antigen in a native form.  Recharged
CC   cells contain exogenous peptides at a higher density than native cells
CC   and are used to stimulate specific cytotoxic T lymphocytes in response to
CC   infections, tumours or autoimmune diseases. This peptide is derived from
CC   ovalbumin and belongs to the allelic restriction Kb.
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 17;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 4
AAW19955
ID   AAW19955 standard; Peptide; 8 AA.
XX
AC   AAW19955;
XX
DT   10-NOV-1997  (first entry)
XX
DE   Chicken OVA-peptide.
XX
KW   Vaccine; immunotherapy; heat shock protein; OVA; cancer;
KW   infectious disease.
XX
OS   Gallus sp.
XX
PN   WO9706821-A1.
XX
PD   27-FEB-1997.
XX
PF   16-AUG-1996;   96WO-US13363.
XX
PR   18-AUG-1995;   95US-0002490.
PR   18-AUG-1995;   95US-0002479.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Hartl FU,  Hoe MH,  Houghton A,  Mayhew M,  Rothman JE;
PI   Takeuchi Y;
XX
DR   WPI; 1997-165035/15.
XX
PT   Compsn. for inducing immune response contg. antigen and heat shock
PT   protein - also new hybrid peptide and related nucleic acid, for
PT   treatment of infectious diseases and tumours
XX
PS   Example 1; Page 17; 58pp; English.
XX
CC   Chicken OVA-peptide (AAW19955) is used in novel hybrid peptides,
CC   OVA-BiP (AAW19956) and BiP-OVA (AAW19957) with heat shock protein (HSP)
CC   BiP binding domain (see also AAW19951).  The hybrid protein is
CC   combined in vitro with a HSP, such as hsp70, to form a complex
CC   that, when administered to a subject, induces an immune response.
CC   Vaccine compositions were prepd. by combining recombinant mouse
CC   hsp70, recombinant human hsp40 and Ova-peptide.  Combinations of
CC   antigen with hsp70 or a mixture of hsp70 and hsp40 were effective
CC   to produce a cytotoxic T lymphocyte response.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 18;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 5
AAW14087
ID   AAW14087 standard; peptide; 8 AA.
XX
AC   AAW14087;
XX
DT   20-OCT-1997  (first entry)
```

```
XX
DE   MHC class I molecule binding OVA peptide.
XX
KW   Major histocompatibility complex; MHC; target; binding; tumour;
KW   cancer; neoplasia; LSTRA; EL-4; identification; detection; screening;
KW   tissue typing; Bcr-abl.
XX
OS   Mus sp.
XX
PN   WO9641188-A1.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;   96WO-US09680.
XX
PR   07-JUN-1995;   95US-0485610.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Cheever MA,  Chen W;
XX
DR   WPI; 1997-108657/10.
XX
PT   Identifying major histocompatibility complex class I binding mols. -
PT   using peptide(s) having a core of 7-14 amino acids with extra amino
PT   acids and a reporter gp. at the N- or C-terminus, useful for tissue
PT   typing
XX
PS   Example 2; Page 19; 41pp; English.
XX
CC   AAW14087-91 are peptides derived from LSTRA and EL-4 tumours of Balb/c
CC   mice. The peptides were tested for MHC specificity to find MHC I
CC   specific peptides. These peptides are useful for tissue typing or for
CC   screening for molecules that interact with MHC class I molecules. MHC
CC   class I molecules can be identified using the peptides and also the
CC   peptides are useful in vaccines against disease and infection e.g.
CC   caused by viruses, bacteria or tumours.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 18;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8

RESULT 6
AAW04642
ID   AAW04642 standard; peptide; 8 AA.
XX
AC   AAW04642;
XX
DT   01-AUG-1997  (first entry)
XX
DE   Ovalbumin-derived activated CD8+ T cells epitope OVA8.
XX
KW   Macrophage; artificial antigen presenting cell; APC; cancer;
KW   tumours; neoplasia; viral infection; retroviral infection;
KW   autoimmune.
XX
OS   Synthetic.
XX
PN   WO9637107-A1.
XX
PD   28-NOV-1996.
XX
PF   22-MAY-1996;   96WO-US07436.
XX
PR   23-MAY-1995;   95US-0447761.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   DeBruijn MLH,  Jackson MR,  Peterson PA;
XX
DR   WPI; 1997-020850/02.
XX
PT   Prodn. of activated CD8+ T cells directed to specific antigen  - can
PT   specifically kill target cells useful to treat, e.g. cancer
XX
PS   Example 1; Page 26; 84pp; English.
XX
CC   The method for the production of activated CD8+ T cells specifically
CC   directed towards a particular antigen involves affixing peptides
CC   corresponding to the particular antigen to an artificial support;
CC   contacting macrophages with the affixed peptides for a time sufficient
CC   for the peptides to be engulfed, and at least a portion of the peptides
CC   to be presented on the surface of the macrophage; and contacting
CC   unprimed CD8+ T cells with the peptide presenting macrophages for a
CC   time sufficient to activate the unprimed CD8+ T cells. The present
CC   sequence represents a peptide designated OVA8 which corresponds to
CC   ovalbumin, a Kb-restricted peptide antigen. This is the optimal
CC   peptide. Small extensions to the optimal peptide affect the affinity
CC   of the peptide for soluble class I molecules in vitro e.g. the addition
CC   of two amino acids to the amino-terminus lowers the affinity to Kb by
CC   76-fold compared to the optimal peptide; addition of two amino acids to
CC   the carboxy-terminus lowers the affinity by 4-fold. The method,
CC   macrophages and artificial antigen presenting cell, having a peptide
CC   corresponding to the particular antigen present on its surface and at
CC   least a portion of an artificial support in its interior, can be used to
CC   treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
CC   infection or autoimmune or autoimmune-type conditions) in patients via
CC   the specific killing of target cells.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 18;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8

RESULT 7
AAW80296
ID   AAW80296 standard; Peptide; 8 AA.
XX
AC   AAW80296;
XX
DT   08-JAN-1999  (first entry)
XX
DE   Amino acids 257-264 of chicken ovalbumin used as an antigen.
XX
KW   Antisense oligonucleotide; antigen processing protein; TAP;
KW   transporter; proteasome; antigen-presenting cell; cancer; infection;
KW   cytotoxic T cell; chicken ovalbumin.
XX
OS   Synthetic.
OS   Gallus sp.
XX
PN   US5831068-A.
XX
PD   03-NOV-1998.
XX
PF   20-AUG-1996;   96US-0700035.
XX
PR   20-AUG-1996;   96US-0700035.
PR   21-AUG-1995;   95US-0517373.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Gilboa E,  Nair SK;
```

```
XX
DR   WPI; 1998-609331/51.
XX
PT   Increasing the presentation of a peptide on a mammalian cell for
PT   production of antigen-presenting cells and stimulation of immune
PT   response - by contacting cells with antigen after inactivating the
PT   protein transporter associated with antigen processing or proteasome
XX
PS   Disclosure; Column 12; 27pp; English.
XX
CC   AAW80296-99 represent peptide antigens used in the course of the
CC   invention. The specification describes a method for increasing the
CC   presentation of a peptide (antigen) on a mammalian cell. The method
CC   comprises inhibiting the activity of a transporter associated with
CC   TAP or proteasome in the cell in vitro before contacting the cell
CC   with the peptide. Antigen-presenting cells produced as above can be used
CC   to stimulate an immune response in vitro or in vivo e.g. to treat or
CC   prevent cancer or infection with a pathogen, e.g. a bacterium or virus.
CC   Cytotoxic T cells produced as above can also be used for therapy.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 19;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 8
AAW70375
ID   AAW70375 standard; Protein; 8 AA.
XX
AC   AAW70375;
XX
DT   18-NOV-1998  (first entry)
XX
DE   Ovalbumin peptide used in the method of the invention.
XX
KW   Ovalbumin; hsp70; heat shock protein; vaccine; tumour therapy.
XX
OS   Synthetic.
OS   Gallus sp.
XX
PN   WO9835705-A1.
XX
PD   20-AUG-1998.
XX
PF   18-FEB-1998;   98WO-US03033.
XX
PR   25-NOV-1997;   97US-0066288.
PR   18-FEB-1997;   97US-0038059.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Young RA;
XX
DR   WPI; 1998-456872/39.
XX
PT   Use of heat shock protein - for delivery of moiety into cells,
PT   useful for vaccination against tumours
XX
PS   Disclosure; Fig 1A; 45pp; English.
XX
CC   The present sequence represents residues 258-276 of the ovalbumin
CC   protein.  The ovalbumin peptide was used in the method of the
CC   invention.  The invention provides a method for delivering a moiety
CC   (e.g. ovalbumin protein) of interest into a cell which involves
CC   contacting the cell with a complex comprising the moiety of interest
CC   covalently linked to a heat shock protein (e.g. hsp70).  The method
CC   is claimed to be useful for providing the efficient delivery into cells
CC   of moieties which are not normally able to enter cells or which enter
CC   cells only to a limited extent.  The method can be useful for delivering
CC   moieties such as proteins, peptides, lipids, glycoproteins, small
CC   organic molecules and other molecules, particularly chemicals and other
CC   molecules which are useful therapeutically or diagnostically.  The
CC   method is claimed to be useful when applied to vaccination regimes,
CC   e.g. for tumour therapy.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 19;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 9
AAW68308
ID   AAW68308 standard; peptide; 8 AA.
XX
AC   AAW68308;
XX
DT   25-MAR-2003  (updated)
DT   14-OCT-1998  (first entry)
XX
DE   MHC binding peptide Ova8.
XX
KW   Antigen; major histocompatibility complex; MHC; lymphocyte; detection;
KW   immobilisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;
KW   viral infection.
XX
OS   Synthetic.
XX
PN   WO9744667-A2.
XX
PD   27-NOV-1997.
XX
PF   21-MAY-1997;   97WO-FR00892.
XX
PR   21-MAY-1996;   96US-0651925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
XX
PI   Langladedemoyen P,  Lone Y,  Kourilsky P,  Abastado J;
XX
DR   WPI; 1998-018653/02.
XX
PT   Detection, purification and elimination of antigen-specific
PT   lymphocytes - for producing cytotoxic T cells for immuno-therapy of
PT   cancers and viral infection
XX
PS   Disclosure; Page 24; 222pp; French.
XX
CC   Peptides AAW68301-W68384 are examples of antigens (Ag) which can be
CC   loaded onto recombinantly produced major histocompatibility complex
CC   (MHC) molecules in a method of detecting antigen-specific lymphocytes.
CC   The MHC-antigen complex is then immobilised on a solid support and a
CC   sample containing cells recognising the MHC-Ag complex may be isolated.
CC   A similar method is used to isolate, purify or eliminate Ag-specific
CC   T-cells or to produce Ag-specific cytotoxic T-cells (CTC).  The method
CC   is also used to detect and quantify tumour-specific T-cells and to
CC   generate CTC for specific killing of tumour cells (solid tumours,
CC   leukaemia or lymphoma) by injection into a human or animal, but also
CC   for treating viral infections.
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 19;  Length 8;
```

```
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 10
AAW68365
ID   AAW68365 standard; peptide; 8 AA.
XX
AC   AAW68365;
XX
DT   25-MAR-2003  (updated)
DT   14-OCT-1998  (first entry)
XX
DE   MHC binding peptide from ovalbumin.
XX
KW   Antigen; major histocompatibility complex; MHC; lymphocyte; detection;
KW   immobilisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;
KW   viral infection.
XX
OS   Synthetic.
XX
PN   WO9744667-A2.
XX
PD   27-NOV-1997.
XX
PF   21-MAY-1997;   97WO-FR00892.
XX
PR   21-MAY-1996;   96US-0651925.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
XX
PI   Langladedemoyen P,  Lone Y,  Kourilsky P,  Abastado J;
XX
DR   WPI; 1998-018653/02.
XX
PT   Detection, purification and elimination of antigen-specific
PT   lymphocytes - for producing cytotoxic T cells for immuno-therapy of
PT   cancers and viral infection
XX
PS   Disclosure; Page 29; 222pp; French.
XX
CC   Peptides AAW68301-W68384 are examples of antigens (Ag) which can be
CC   loaded onto recombinantly produced major histocompatibility complex
CC   (MHC) molecules in a method of detecting antigen-specific lymphocytes.
CC   The MHC-antigen complex is then immobilised on a solid support and a
CC   sample containing cells recognising the MHC-Ag complex may be isolated.
CC   This peptide is derived from amino acids 258-276 of ovalbumin.  A
CC   similar method is used to isolate, purify or eliminate Ag-specific
CC   T-cells or to produce Ag-specific cytotoxic T-cells (CTC).  The method is
CC   also used to detect and quantify tumour-specific T-cells and to generate
CC   CTC for specific killing of tumour cells (solid tumours, leukaemia or
CC   lymphoma) by injection into a human or animal, but also for treating
CC   viral infections.
CC   (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 19;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 11
AAW60700
ID   AAW60700 standard; peptide; 8 AA.
XX
AC   AAW60700;
XX
DT   22-SEP-1998  (first entry)
XX
DE   Ovalbumin peptide Tctl peptide epitope (residues 257-264).
XX
KW   Immunisation; target antigen; epitope; inoculation; infant mammal;
KW   viral antigen; depressed humoral response; respiratory syncytial virus;
KW   rotavirus; measles virus; human immunodeficiency virus; hepatitis virus;
KW   herpes simplex virus; influenza virus; Streptococcus pneumoniae;
KW   Hemophilus influenzae; Neisseria meningitidis; Staphylococcus aureus;
KW   protozoan antigen; malaria.
XX
OS   Unidentified.
XX
PN   WO9822145-A1.
XX
PD   28-MAY-1998.
XX
PF   21-NOV-1997;   97WO-US21687.
XX
PR   22-NOV-1996;   96US-0755034.
XX
PA   (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
PI   Bona C,  Bot A;
XX
DR   WPI; 1998-312182/27.
XX
PT   Immunisation of infant mammals - by inoculating the mammal with a
PT   nucleic acid encoding a relevant epitope of a target antigen
XX
PS   Disclosure; Page 10; 83pp; English.
XX
CC   Sequence shown in AAW60683 to AAW60700 are epitope sequences of various
CC   viral antigens used to exemplify the method of invention of immunising
CC   an infant mammal against a target antigen. The method comprises
CC   inoculating the mammal with a nucleic acid encoding a relevant epitope
CC   of a target antigen in a carrier, such that the relevant epitope is
CC   expressed in the infant mammal. The genetic immunisation of infant
CC   mammals can give rise to effective cellular (including the induction of
CC   cytotoxic T lymphocytes) and humoral immune responses against the target
CC   antigen. The methods are particularly used for treating infants with
CC   depressed humoral responses, that have high-zone tolerances against the
CC   target antigens or have a Th2 biased immune response. The target antigen
CC   may be a viral antigen, e.g. a respiratory syncytial virus antigen, a
CC   rotavirus antigen, a measles virus antigen, a human immunodeficiency
CC   virus antigen, a hepatitis virus antigen, a hepatitis B virus antigen, a
CC   herpes simplex virus antigen or an influenza virus antigen, a bacterial
CC   antigen e.g. Streptococcus pneumoniae antigen, Hemophilus influenzae
CC   antigen, Neisseria meningitidis antigen, Staphylococcus aureus antigen
CC   or a protozoan antigen such as a malaria antigen.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 19;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 12
AAW54265
ID   AAW54265 standard; peptide; 8 AA.
XX
AC   AAW54265;
XX
DT   30-JUL-1998  (first entry)
```

```
XX
DE   Ovalbumin peptide OVA.
XX
KW   Pep-MHC complex; cytotoxic; T cell; cancer; ovarian; brain;
KW   major histocompatibility complex.
XX
OS   Synthetic.
XX
PN   WO9807441-A1.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-US14814.
XX
PR   23-AUG-1996;   96US-0023437.
XX
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
PA   (UNII ) UNIV ILLINOIS FOUND.
XX
PI   Eisen HN,  Kranz DM;
XX
DR   WPI; 1998-168897/15.
XX
PT   Composition for targeting an allo-reactive response to specific
PT   cells - comprises MHC-peptide complex bound to specific targeting
PT   agent, the complex having at least one allogeneic component,
PT   specifically for treating ovarian and brain cancers
XX
PS   Examples; Page 26; 45pp; English.
XX
CC   Peptide OVA derived from ovalbumin was used in the production of soluble
CC   pep-MHC complexes.  The pep-MHC (peptide-major histocompatibility
CC   complex) complexes, including both MHC heavy and light chains, bound to a
CC   specific targeting molecule can be used to target a cytotoxic T cell
CC   response to specific cells.  This is particularly useful for targeting
CC   cancer cells, specifically of the ovary or brain.  The method can also
CC   be used to eliminate an entire cell type for example during bone marrow
CC   therapy.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 19;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                ||||||||
Db          1 SIINFEKL 8


RESULT 13
AAW52838
ID   AAW52838 standard; peptide; 8 AA.
XX
AC   AAW52838;
XX
DT   24-JUN-1998  (first entry)
XX
DE   Ovalbumin artificial target antigen.
XX
KW   Ovalbumin; antigen; ATA; cytotoxic T lymphocyte; CTL; tumour; prevention.
XX
OS   Gallus sp.
XX
PN   WO9800163-A1.
XX
PD   08-JAN-1998.
XX
PF   18-JUN-1997;   97WO-US10195.
XX
PR   28-JUN-1996;   96US-0675332.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Falo LD,  Rock KL;
XX
DR   WPI; 1998-086733/08.
XX
PT   Inducing anti-tumour cytotoxic T lymphocytes  - by cross-priming
PT   with artificial antigen, then immunisation with tumour cells
PT   modified in vitro to express the same antigen, does not require
PT   characterisation of tumour-specific antigens
XX
PS   Example; Page 21; 44pp; English.
XX
CC   Ovalbumin can be used as an artificial target antigen (ATA) to
CC   promote a cytotoxic T lymphocyte mediated response in mammals.  Tumour
CC   cells from the host can be engineered to include ATA and therefore
CC   induce anti-tumour cytotoxic T Lymphocytes. This method can be used for
CC   the treatment and prevention of a wide range of tumours even when
CC   the tumour is inaccessible or where metastases are being targeted.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 19;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                ||||||||
Db          1 SIINFEKL 8


RESULT 14
AAY42307
ID   AAY42307 standard; peptide; 8 AA.
XX
AC   AAY42307;
XX
DT   06-DEC-1999  (first entry)
XX
DE   Ovalbumin-derived peptide antigen.
XX
KW   Immunity; human leukocyte antigen; HLA; MHC; antigen;
KW   major histocompatibility complex; presentation; solubility;
KW   dendritic cell.
XX
OS   Synthetic.
XX
PN   WO9947646-A1.
XX
PD   23-SEP-1999.
XX
PF   19-MAR-1999;   99WO-US06627.
XX
PR   20-MAR-1998;   98US-0078832.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (BIOP-) BIOPOLO SCRL.
XX
PI   Rescigno M,  Girolomoni G,  Corinti S,  Ricciardi-Castagnoli P;
XX
DR   WPI; 1999-571834/48.
XX
PT   Preparation of dendritic cells which present antigens, used for
PT   stimulating an immune response by immunocompetent cells  -
XX
PS   Example 6; Page 13; 43pp; English.
XX
CC   This sequence represents an ovalbumin-derived peptide antigen which
CC   is presented on major histocompatibility complex (MHC) Class I molecules
CC   of dendritic cells via the use of a novel process to improve soluble
CC   protein antigen presentation. This process uses dendritic cells which
CC   have internalised bacterial cells recombinantly expressing ovalbumin. The
CC   dendritic cells which have internalised the bacteria can be used to
```

```
CC   generate an immune response which is stronger than the response which
CC   would be generated when soluble antigen alone is used. The dendritic
CC   cells which result from the internalisation are characterised by MHC
CC   Class I and Class II molecules which have much longer half lives and
CC   greater stability than comparable dendritic cells which have not
CC   internalised such bacteria. The dendritic cells can be contacted with
CC   immunocompetent cells for stimulating an immune response. The dendritic
CC   cells can also be used to stimulate maturation of an immature dendritic
CC   cell.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 20;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 15
AAY16871
ID   AAY16871 standard; peptide; 8 AA.
XX
AC   AAY16871;
XX
DT   20-JUL-1999  (first entry)
XX
DE   OVA peptide sequence.
XX
KW   Conjugate peptide; heat shock protein; hsp; phage display library; virus;
KW   surface protein; tethering peptide; chaperone process; cytokine; cancer;
KW   neoplastic disease; infectious disease; bacterium; immune system; fungus;
KW   acquired immune deficiency; autoimmune disease.
XX
OS   Synthetic.
XX
PN   WO9922761-A1.
XX
PD   14-MAY-1999.
XX
PF   22-OCT-1998;   98WO-US22335.
XX
PR   31-OCT-1997;   97US-0961707.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Hartl U,  Hoe MH,  Houghton A,  Mayhew M,  Moroi Y;
PI   Ouerfelli O,  Rothman JE;
XX
DR   WPI; 1999-313177/26.
XX
PT   Identifying peptides which bind heat shock proteins
XX
PS   Examples; Page 49; 155pp; English.
XX
CC   The invention relates to conjugate peptides engineered to noncovalently
CC   bind to heat shock proteins (hsp). A method of identifying a hsp binding
CC   peptide comprises (a) contacting a phage display library having
CC   bacteriophage expressing, in a surface protein, inserted peptides with a
CC   hsp target, and bound to a benzoquinone ansamycin antibiotic (BAA), in a
CC   physiologic binding buffer; (b) isolating a phage binding to the hsp
CC   target; and (c) identifying the inserted peptide expressed. The peptides
CC   which bind to a hsp can be used as tethering peptides for a hsp which may
CC   serve as an accessory in a chaperone process and/or may comprise a
CC   cytokine. They can also be coupled to antigens to induce an immune
CC   response. Such compositions can be used for treating neoplastic disease,
CC   e.g. cancers, infectious diseases, e.g. diseases caused by a bacterium,
CC   virus, protozoan, mycoplasma, fungus, yeast, parasite or prion, or a
CC   disease of the immune system, e.g. acquired immune deficiencies or
CC   autoimmune diseases.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 20;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 16
AAY03780
ID   AAY03780 standard; peptide; 8 AA.
XX
AC   AAY03780;
XX
DT   23-JUN-1999  (first entry)
XX
DE   Ovalbumin peptide fragment (residues 257-264).
XX
KW   Dendritic cell-derived factor; proliferation; interferon gamma; IFNgamma;
KW   T cell; granulocyte macrophage colony-stimulating factor; GM-CSF;
KW   T cell stimulatory factor; lymphocyte; interleukin-2; lipopolysaccharide;
KW   autoimmune response; inflammatory; ovalbumin.
XX
OS   Synthetic.
OS   Gallus gallus.
XX
PN   WO9918909-A2.
XX
PD   22-APR-1999.
XX
PF   14-OCT-1998;   98WO-US21614.
XX
PR   14-OCT-1997;   97US-0062405.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Dunn A,  Marino MW,  Noguchi Y,  Old LJ,  Wada H;
XX
DR   WPI; 1999-277418/23.
XX
PT   Dendritic cell and T cell derived factors for regulation of T cell
PT   proliferation and interferon gamma production
XX
PS   Examples; Page 37; 68pp; English.
XX
CC   The invention relates to a dendritic cell-derived factor that restores
CC   proliferation and interferon gamma (IFNgamma) production to T cells from
CC   granulocyte macrophage colony-stimulating factor (GM-CSF). The dendritic
CC   cell-derived factor which is a T cell stimulatory factor modulates the
CC   effect of GM-CSF on production of IFNgamma by lymphocytes and the
CC   response of GM-CSF -/- T cells to interleukin-2, and corrects the
CC   lipopolysaccharide-induced defect in IFNgamma production. The dendritic
CC   cell-derived factor is used in vivo to increase proliferation of T cells
CC   and/or IFNgamma production, e.g. during immunisation to increase the
CC   response to an antigen. Agents with T cell derived factor activity are
CC   used to decrease production of IFNgamma, e.g. to reduce autoimmune
CC   responses and inflammatory reactions. Agents that bind to the factors,
CC   e.g. antisense sequences, are used to treat excessive/inadequate T cell
CC   proliferation and IFNgamma production. The dendritic cell-derived factor
CC   and T cell factor (or their fragments) can also be used to raise
CC   antibodies and as components in immunoassay and diagnostic systems, also
CC   for in vitro studies. The present sequence represents an ovalbumin
CC   peptide fragment used in the course of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 20;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 17
AAW99480
ID   AAW99480 standard; peptide; 8 AA.
XX
AC   AAW99480;
XX
DT   08-JUN-1999  (first entry)
XX
DE   Ovalbumin-derived peptide OVA.
XX
KW   Matrix protein; conjugate; mutant; major histocompatibility complex; MHC;
KW   class I molecule; beta-2-microglobulin; stimulation; immunity; tumour.
XX
OS   Synthetic.
XX
PN   WO9911775-A1.
XX
PD   11-MAR-1999.
XX
PF   20-AUG-1998;   98WO-US17308.
XX
PR   29-AUG-1997;   97US-0920413.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Garboczi DN,  Walker J;
XX
DR   WPI; 1999-205182/17.
XX
PT   Method for conjugating a mutant major histocompatibility complex
PT   class I molecule and a compound - useful for stimulating immunity in
PT   an individual, and eradicating undesired cells, especially tumours
XX
PS   Example III-4; Page 22; 37pp; English.
XX
CC   The invention relates to the preparation of a conjugate of a mutant major
CC   histocompatibility complex (MHC) class I molecule (containing a Tyr67Cys
CC   amino acid substitution in the beta2-microglobulin subunit) and a
CC   compound.  This sequence corresponds to the ovalbumin-derived peptide
CC   OVA.  The peptide is used as a control to prepare a hybrid peptide-MHC
CC   class I tetramer in which a mutant human beta2-microglobulin subunit
CC   binds a mouse MHC class I molecule.  The conjugates are useful for
CC   stimulating immunity in an individual, and eradicating undesired cells
CC   (e.g. tumours).
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 20;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 18
AAW67585
ID   AAW67585 standard; peptide; 8 AA.
XX
AC   AAW67585;
XX
DT   02-MAR-1999  (first entry)
XX
DE   T-cell activation peptide #3.
XX
KW   Activated T helper cell; CD4+; cytotoxic T cell; CD8+; liposome; epitope;
KW   peripheral blood lymphocyte; antigen-presenting cell; APC; virus; tumour;
KW   bacterium; parasite; cytokine; vaccine; cancer; malaria; HIV; hepatitis;
KW   tuberculosis.
XX
OS   Synthetic.
XX
PN   WO9850527-A1.
XX
PD   12-NOV-1998.
XX
PF   07-MAY-1998;   98WO-US09288.
XX
PR   08-MAY-1997;   97US-0045949.
XX
PA   (BIOM-) BIOMIRA INC.
XX
PI   Agrawal B,  Krantz MJ,  Longenecker BM,  Reddish MA;
XX
DR   WPI; 1999-034715/03.
XX
PT   Method of activation of T cells  - by exposure to antigen-presenting
PT   cells loaded with antigen in liposome, used for, e.g. treating
PT   cancer and microbial infections
XX
PS   Disclosure; Page 6; 75pp; English.
XX
CC   Peptides AAW67583-W67611 are used to produce activated T helper (CD4+)
CC   and cytotoxic (CD8+) T cells. The activated T cells are produced by
CC   treating peripheral blood lymphocytes with liposome-encapsulated peptide
CC   antigen to generate Ag-loaded antigen-presenting cells (APC), contacting
CC   naive or anergic T-cells with these APC, and isolating the resulting
CC   activated T-cells.  The cells are specific for a particular antigen,
CC   particularly one derived from a tumour, but also those from viruses,
CC   bacteria and other parasites. It can also be used to identify antigens
CC   and epitopes able to generate an Ag-specific T-cell response (by
CC   assessing proliferation and cytokine release).  Also the Ag-loaded APC
CC   can be used as cellular vaccines for treating cancer (claimed) or other
CC   diseases (e.g. malaria, human immune deficiency virus infection,
CC   hepatitis, tuberculosis).  The activated T-cells can be used to treat the
CC   same conditions by adoptive T-cell transfer therapy.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 20;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 19
AAB13763
ID   AAB13763 standard; peptide; 8 AA.
XX
AC   AAB13763;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Peptide fragment from ovalbumin OVA.
XX
KW   T-cell; immune response; antigen; epitope; B7 family molecule;
KW   Leukocyte function-associated antigen-3; LFA-3;
KW   Intercellular adhesion molecule-1; ICAM-1; vaccine; immunotherapy;
KW   colon polyp; Crohn's disease; ulcerative colitis; breast lesion;
KW   tumour; ovalbumin.
XX
OS   Unidentified.
XX
PN   WO200034494-A1.
XX
PD   15-JUN-2000.
```

```
XX
PF   12-NOV-1999;   99WO-US26866.
XX
PR   09-DEC-1998;   98US-0111582.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA   (THER-) THERION BIOLOGICS CORP.
XX
PI   Schlom J,  Hodge J,  Panicali D;
XX
DR   WPI; 2000-431307/37.
XX
PT   Novel recombinant vector useful as immunogens and vaccines for
PT   stimulating and enhancing immunological responses to target cells and
PT   antigens expresses multiple co-stimulatory molecules such as B7-1,
PT   LFA-3, ICAM-1  -
XX
PS   Example 31; Page 80; 188pp; English.
XX
CC   Costimulatory molecules have important roles in T-cell activation and
CC   therefore the immune response. The present invention relates to
CC   recombinant vectors which comprise of foreign nucleic acid sequences
CC   encoding at least three costimulatory molecules: a B7 family molecule,
CC   Leukocyte function-associated antigen-3 (LFA-3, human CD58) and
CC   Intercellular adhesion molecule-1 (ICAM-1, CD54) and optionally a foreign
CC   gene encoding a target antigen or immunological epitope. The present
CC   sequence is one such target antigen used in the present invention. The
CC   present sequence is a tumour-associated antigen. The vector of the
CC   present invention would be useful for providing an enhanced immune
CC   response to the present target antigen. The vector of the present
CC   invention may therefore be useful in immunotherapy for treating or
CC   preventing diseases caused by viruses, bacteria, protozoans, parasites,
CC   premalignant cells and tumour cells. The recombinant vector can be used
CC   to treat or prevent preneoplastic or hyperplastic states such as colon
CC   polyps, Crohn's disease, ulcerative colitis and breast lesions.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 20
AAB29465
ID   AAB29465 standard; peptide; 8 AA.
XX
AC   AAB29465;
XX
DT   09-FEB-2001  (first entry)
XX
DE   Negative control peptide used in a cytotoxic T cell assay.
XX
KW   Telomerase antigen variant; HLA-A2-binding; class I MHC;
KW   human leukocyte antigen; major histocompatibility complex;
KW   cytotoxic T-cell response; antigen-presenting cell; APC;
KW   telomerase-expressing cell; cancer; anticancer vaccine.
XX
OS   Synthetic.
XX
PN   WO200061766-A2.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-IB00610.
XX
PR   09-APR-1999;   99US-0128539.
XX
PA   (BIOM-) BIOMIRA INC.
XX
PI   Agrawal B,  Longenecker BM;
XX
DR   WPI; 2000-679493/66.
XX
PT   New telomerase-specific T-cell antigens useful for generating T-cell
PT   responses against telomerases and for producing vaccines for treating
PT   or preventing cancer by in vivo or ex vivo techniques  -
XX
PS   Example 1; Page 23; 34pp; English.
XX
CC   The invention relates to a human telomerase peptide antigen (AAB29461)
CC   which binds to a class I HLA (human leukocyte antigen, MHC, major
CC   histocompatibility complex), and to conservatively substituted variants
CC   thereof. The invention also relates to a vaccine comprising a telomerase
CC   antigen or antigen variant, a nucleotide encoding a telomerase antigen
CC   or variant, and a method of producing telomerase-primed antigen-
CC   presenting cell (APC) comprising contacting an APC with a composition
CC   containing a telomerase antigen or variant. The telomerase antigens or
CC   vaccine compositions are useful for inducing a cytotoxic T-cell immune
CC   response against telomerase and hence against telomerase-expressing
CC   cells (i.e., cancer cells. Additionally, the telomerase antigen-primed
CC   APC may be coadministered with interleukin-2 for cancer treatment or
CC   prevention. The present sequence represents a peptide used in the
CC   exemplification of the invention in an assay of the cytotoxic activity of
CC   T-cells.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 21
AAB26484
ID   AAB26484 standard; peptide; 8 AA.
XX
AC   AAB26484;
XX
DT   16-JAN-2001  (first entry)
XX
DE   Tumour associated OVA peptide.
XX
KW   Immune response; vaccine; cancer; infection; tumour; OVA.
XX
OS   Unidentified.
XX
PN   WO200050080-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-US04565.
XX
PR   26-FEB-1999;   99US-0261473.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Gilboa E,  Nair SK,  Nicchitta CV;
XX
DR   WPI; 2000-558368/51.
XX
PT   Eliciting immune response in vertebrate for prevention and treatment of
PT   cancer and infectious diseases involves administering purified complex
PT   comprising calreticulin bound to an antigenic molecule  -
XX
PS   Disclosure; Page 7; 82pp; English.
XX
CC   The present invention relates to a method of eliciting an immune
```

```
CC   response by administering a composition of a purified complex of
CC   calreticulin bound to an antigenic molecule. The present invention is
CC   useful for prevention and treatment of cancer and infectious disease
CC   in a vertebrate especially of humans. The present sequence is the
CC   tumour antigenic peptide OVA which was used in the invention.
XX
SQ   Sequence   8 AA;

  Query Match            100.0%;  Score 38;  DB 21;  Length 8;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 22
AAB13785
ID   AAB13785 standard; peptide; 8 AA.
XX
AC   AAB13785;
XX
DT   10-NOV-2000  (first entry)
XX
DE   OVA-derived peptide.
XX
KW   Chicken; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW   major histocompatibility complex class 1; MHC class 1; antigen; tumour;
KW   prostate; breast; multiple myeloma; OVA peptide.
XX
OS   Gallus domesticus.
XX
PN   WO200035949-A1.
XX
PD   22-JUN-2000.
XX
PF   14-DEC-1999;   99WO-US29724.
XX
PR   14-DEC-1998;   98US-0112324.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Hakim I,  Vidovic D;
XX
DR   WPI; 2000-442365/38.
XX
PT   Antigens modified by the covalent addition of a peptide that
PT   facilitates entry into antigen presenting cells, useful for producing
PT   compositions for immunizing against tumors and pathogens -
XX
PS   Example 1; Page 10; 34pp; English.
XX
CC   The present invention relates to compositions of modified soluble protein
CC   antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC   response i.e. a major histocompatability complex (MHC) class I molecule
CC   response. The protein antigen is modified by the covalent addition of a
CC   peptide sequence which facilitate entry of the antigen into antigen
CC   presenting cells (APCs). The present sequence is a peptide derived from
CC   the chicken antigen OVA. This peptide was used to prepare the modified
CC   antigens. The modified antigen composition may be used for immunising
CC   against, or treating a tumour e.g. prostate and breast carcinoma or
CC   multiple myeloma, or pathogen in mammals.
XX
SQ   Sequence   8 AA;

  Query Match            100.0%;  Score 38;  DB 21;  Length 8;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 23
AAY68311
ID   AAY68311 standard; Peptide; 8 AA.
XX
AC   AAY68311;
XX
DT   13-APR-2000  (first entry)
XX
DE   Altered MHC determinant binding peptide SEQ ID NO:143.
XX
KW   MHC class I; major histocompatibility complex; microglobulin; antigen;
KW   immune response; immunisation; AIDS; multiple sclerosis; toxic shock;
KW   cancer; lupus erythematosus; snake bite; cytostatic; antiviral;
KW   immunomodulatory; dermatological; immunosuppressive; antiinflammatory;
KW   neuroprotective.
XX
OS   Unidentified.
XX
PN   US6011146-A.
XX
PD   04-JAN-2000.
XX
PF   07-JUN-1995;   95US-0481985.
XX
PR   15-NOV-1991;   91US-0792473.
PR   05-DEC-1991;   91US-0801818.
XX
PA   (INSP ) INST PASTEUR.
PA   (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI   Kourilsky P,  Mottez E,  Abastado J;
XX
DR   WPI; 2000-125951/11.
XX
PT   New recombinant DNA encoding covalently linked form of major
PT   histocompatibility complex Class I determinant, used for immune system
PT   stimulation, e.g. for treating cancer  -
XX
PS   Disclosure; Column 12; 88pp; English.
XX
CC   The present invention describes a recombinant DNA molecule (I)
CC   containing a sequence (Ia) that encodes an altered MHC (major
CC   histocompatibility complex ) Class I determinant (II) comprises a
CC   polypeptide with alpha1, alpha2, alpha3 and beta2-microglobulin
CC   domains, in which alpha3 and beta2 are covalently linked, thorough C-
CC   and N-termini respectively, via a nucleotide spacer sequence encoding a
CC   polypeptide. (II) includes an antigen-binding site and when (II) and
CC   the antigen are associated they are recognized by a mammalian T cell
CC   receptor (TCR). (I) are used to produce (II) which are used to study
CC   functional interactions between the various MHC domains. They can also
CC   be used to modulate (in vivo or in vitro) the immune system by inducing
CC   an effector response (cytotoxicity, antibody synthesis, phagocytosis)
CC   of immune system cells, typically for treating, or immunising against;
CC   cancer, acquired immune deficiency syndrome, lupus erythematosus,
CC   multiple sclerosis, toxic shock and snake bite, but also for selective
CC   destruction of autoreactive cells, diagnostically to assay T cell
CC   receptors and to raise specific antibodies (useful for diagnosis,
CC   therapy, studying MHC-associated cellular processes and for affinity
CC   purification). AAZ57558 and AAY68186 to AAY68316 are sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   8 AA;

  Query Match            100.0%;  Score 38;  DB 21;  Length 8;
  Best Local Similarity  100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8
```

```
RESULT 24
AAY59401
ID   AAY59401 standard; Protein; 8 AA.
XX
AC   AAY59401;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Ovalbumin protein fragment.
XX
KW   Ovalbumin; chicken; nonvirulent bacterium; cytolysin;
KW   diagnosis; gene therapy; polypeptide delivery; gene delivery.
XX
OS   Gallus sp.
XX
PN   US6004815-A.
XX
PD   21-DEC-1999.
XX
PF   13-AUG-1998;   98US-0133914.
XX
PR   13-AUG-1998;   98US-0133914.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Portnoy DA,  Higgins DE;
XX
DR   WPI; 2000-072064/06.
XX
PT   Nonvirulent bacterium useful for the intracellular delivery of agents
PT   to eukaryotic cells -
XX
PS   Example; Column 12; 14pp; English.
XX
CC   This sequence represents a fragment of the chicken ovalbumin.
CC   The invention relates to a nonvirulent bacterium (I), comprising
CC   a first gene encoding a nonsecreted foreign functional cytolysin operably
CC   linked to a heterologous promoter which expresses cytolysin in the
CC   bacterium and a second gene encoding a different foreign agent. The
CC   bacterium can be used to deliver a wide variety of foreign agents to
CC   eukaryotic cells for applications such as diagnosis, therapy including
CC   prophylactic treatments such as immunisations and gene therapy especially
CC   for single gene disorders suitable for localised treatment and
CC   biosynthesis. They can be used to deliver antigenic polypeptides
CC   presented in association with major histocompatibility (MHC) proteins to
CC   antigen-presenting cells. The bacterium can be engineered to deliver
CC   libraries of agents for screening. The foreign agent can be delivered to
CC   any target cell capable of carrying out endocytosis of the subject
CC   microbe, in particular epithelial cells, endothelial cells, muscle cells,
CC   liver cells, pancreatic cells, neural cells, fibroblasts, tumour cells
CC   and leukocytes. The foreign agent introduced into the cell is a nucleic
CC   acid and/or protein which is bioactive in and therapeutic to the target
CC   eukaryote. Bacterium mediated delivery of protein does not require
CC   protein purification and high levels of protein can be delivered to the
CC   cytosol of virtually all cells in culture. The process is rapid and
CC   efficient and large enzymatically active proteins can be delivered.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8

RESULT 25
AAY52564
ID   AAY52564 standard; peptide; 8 AA.
XX
AC   AAY52564;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Murine ovalbumin MHC class I epitope.
XX
KW   Chimeric; ovalbumin; pan DR epitope; expression vector;
KW   promoter; major histocompatibility complex; MHC; targetting; peptide;
KW   epitope; antigen; presentation; class I; cytosolic pathway;
KW   endoplasmic reticulum; class II; extracellular antigen;
KW   endocytic pathway; helper T lymphocyte; HTL; universal epitope;
KW   cytotoxic T lymphocyte; CTL; immune response; immunogenicity; assay;
KW   vaccine; immunity; infection; pathogen; virus; HIV; HBV; HCV;
KW   hepatitis B; hepatitis C; bacterium; protozoan; tumour cell;
KW   autoimmune disease; activation; antiviral; antimalarial;
KW   immunoprotective.
XX
OS   Synthetic.
OS   Mus sp.
XX
PN   WO9958658-A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999;   99WO-US10646.
XX
PR   13-MAY-1998;   98US-0078904.
PR   15-MAY-1998;   98US-0085751.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Fikes JD,  Hermanson GG,  Sette A,  Ishioka GY,  Livingston B;
PI   Chesnut RW;
XX
DR   WPI; 2000-039103/03.
DR   N-PSDB; AAZ38683.
XX
PT   Expression vectors encoding major histocompatibility targeting
PT   sequence, used as, e.g. tumor vaccines -
XX
PS   Example 1; Page 45; 130pp; English.
XX
CC   This sequence represents a murine ovalbumin MHC class I epitope
CC   encoded by the AOS minigene insert of the expression vector pMIN.0
CC   (AAZ38634). This insert encodes several MHC class I epitopes,
CC   including this sequence, plus DNA encoding the universal MHC
CC   class II (helper T) epitope, pan DR epitope (PADRE), and was used
CC   in an exemplification of the present invention. The invention
CC   relates to a novel expression vector comprising a promoter operably
CC   linked to a fusion gene encoding a major histocompatibility complex
CC   (MHC) targetting sequence, and two or more heterologous peptide
CC   epitopes. The MHC targetting sequence may be a class I targetting
CC   sequence, which directs an MHC class I epitope to a cytosolic pathway or
CC   to the endoplasmic reticulum, or an MHC class II targetting sequence,
CC   which directs extracellular antigens to enter the endocytic pathway to
CC   be processed into antigen peptides for presentation on MHC class II
CC   molecules. The heterologous epitopes may comprise either helper T
CC   lymphocyte (HTL) epitopes, or a cytotoxic T lymphocyte (CTL) epitope and
CC   a universal HTL epitope such as a pan DR epitope (PADRE). The vectors
CC   are useful for stimulating an immune response in vivo, as well as for use
CC   in assaying the human immunogenicity of a human T cell peptide epitope
CC   in vivo in a non-human mammal. They provide a nucleic acid vaccine for
CC   enhancing immunity against infectious pathogens, such as viruses (e.g.,
CC   HIV, hepatitis B (HBV and hepatitis C (HCV)) bacteria, protozoa (e.g.,
CC   Plasmodium falciparum, the cause of malaria) and also tumour cells and
CC   autoimmune diseases. Universal MHC class II epitopes are advantageously
CC   combined with other MHC class I and class II epitopes to increase the
CC   number of cells that are activated in response to a given antigen and
CC   provide a broader population coverage of MHC-reactive alleles.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 8;
```

```
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 26
AAY52965
ID   AAY52965 standard; Peptide; 8 AA.
XX
AC   AAY52965;
XX
DT   14-FEB-2000  (first entry)
XX
DE   Altered MHC determinant binding peptide SEQ ID NO:143.
XX
KW   Major histocompatibility complex; MHC class I; MHC class II; antigen;
KW   immune response; diagnosis; antibody; immunisation; autoimmune disease;
KW   acquired immune deficiency syndrome; AIDS; cytostatic; dermatological;
KW   anti-inflammatory; neuroprotective; immunosuppressive; antithyroid;
KW   vaccine; lupus erythematosus; multiple sclerosis; thyroiditis;
KW   toxic shock; tumour; snakebite.
XX
OS   Synthetic.
OS   Synthetic.
XX
PN   US5976551-A.
XX
PD   02-NOV-1999.
XX
PF   07-JUN-1995;   95US-0484905.
XX
PR   05-DEC-1991;   91US-0801818.
PR   15-NOV-1991;   91US-0792473.
XX
PA   (INSP ) INST PASTEUR.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Kourilsky P,  Mottez E,  Abastado J;
XX
DR   WPI; 2000-037081/03.
XX
PT   Composition containing an antigen and altered major histocompatibility
PT   Class II determinant, used to immunize against autoimmune diseases,
PT   e.g. acquired immune deficiency syndrome  -
XX
PS   Claim 8; Column 13; 96pp; English.
XX
CC   The present invention describes a composition capable of eliciting
CC   anti-major histocompatibility (MHC) antibodies. The composition
CC   comprises an antigen associated with an altered MHC Class II determinant
CC   (I) comprising alpha1, alpha2, beta1 and beta2 polypeptide domains
CC   encoded by a mammalian MHC Class II locus covalently linked to form a
CC   polypeptide (I) containing beta2, alpha2, alpha1 and beta1 domains in
CC   sequence. The resulting Antigen-MHC complex is recognizable by the T cell
CC   receptor. The compositions are used for immunisation against, or
CC   treatment of, a wide range of autoimmune diseases, e.g. acquired immune
CC   deficiency syndrome (AIDS), lupus erythematosus, multiple sclerosis,
CC   thyroiditis, toxic shock, tumour and snakebite, depending on the nature
CC   of antigen. (I) is also used to analyse functional interactions between
CC   the various domains and for targeting lymphocyte receptors. Antibodies
CC   against (I) are produced by usual methods of immunisation or cell fusion,
CC   and may be humanised by standard methods. These antibodies are useful for
CC   diagnosis (detection or purification of MHC gene products), therapy
CC   (neutralising MHC on cell surfaces) and in the study of MHC and cellular
CC   processes. AAZ33240 to AAZ33242 and AAY52840 to AAY52970 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 27
AAG77871
ID   AAG77871 standard; Protein; 8 AA.
XX
AC   AAG77871;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Ovalbumin 257-264 peptide.
XX
KW   Ovalbumin 257-264 peptide; major histocompatibility complex;
KW   MHC-peptide complex; MHC; human; MHC class I alpha chain;
KW   beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;
KW   vaccine; immune response modulation; hyperproliferative disorder;
KW   neoplasm; hypergammaglobulinaemia; viral infection; hepatitis;
KW   meningitis; bacterial infection; tuberculosis; gingivitis;
KW   parasitic infection; autoimmune disease; Hashimoto's disease;
KW   Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;
KW   graft-versus-host disease; GVHD; antigenic peptide.
XX
OS   Aves.
XX
PN   WO200178768-A2.
XX
PD   25-OCT-2001.
XX
PF   12-APR-2001; 2001WO-US11912.
XX
PR   12-APR-2000; 2000US-196472P.
XX
PA   (UYRP ) UNIV ROCHESTER.
XX
PI   Zauderer M,  Smith ES;
XX
DR   WPI; 2001-602927/68.
XX
PT   Novel compound comprising major histocompatibility complex-peptide
PT   complexes, used to modulate immune responses -
XX
PS   Example 19; Page 101; 166pp; English.
XX
CC   The invention comprises a compound which contains one or more major
CC   histocompatibility complex (MHC)-peptide complexes, and an antibody
CC   specific for a cell surface marker. The complexes comprise an MHC class
CC   I alpha chain, a beta-2 microglobulin molecule and an antigenic peptide
CC   bound in the MHC groove. Alternatively, the complexes may comprise an MHC
CC   class II alpha chain, an MHC class II beta chain, and an antigenic
CC   peptide bound in the MHC groove. The complexes are linked to the carboxyl
CC   terminus of the antibody. The compounds of the invention can be used as a
CC   vaccine to modulate an immune response. The compounds of the invention
CC   are useful for treating: hyperproliferative disorders (e.g. neoplasms and
CC   hypergammaglobulinaemia); viral infections (e.g. hepatitis and
CC   meningitis); bacterial infections (e.g. tuberculosis and gingivitis);
CC   parasitic infections; autoimmune diseases (e.g. Hashimoto's disease,
CC   Graves' disease and rheumatoid arthritis); allergic reactions/conditions
CC   (e.g. asthma). The compounds of the invention may also be used in the
CC   treatment of organ rejection or graft-versus-host disease (GVHD). The
CC   present amino acid sequence represents the ovalbumin 257-264 peptide,
CC   which was used as an antigenic peptide in an example of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 28
AAM52562
ID   AAM52562 standard; Peptide; 8 AA.
XX
AC   AAM52562;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Cytotoxic T-cell epitope for ovalbumin.
XX
KW   Cell death; toxic gene; tumour suppressor; ovalbumin;
KW   Cytotoxic T-cell epitope.
XX
OS   Unidentified.
XX
PN   WO200172995-A2.
XX
PD   04-OCT-2001.
XX
PF   28-MAR-2001; 2001WO-US09953.
XX
PR   28-MAR-2000; 2000US-0192586.
PR   10-MAY-2000; 2000US-0203343.
PR   23-JAN-2001; 2001US-0263226.
PR   27-FEB-2001; 2001US-0271426.
XX
PA   (UYRP ) UNIV ROCHESTER.
XX
PI   Zauderer M,  Smith ES;
XX
DR   WPI; 2001-570897/64.
XX
PT   Selecting target polynucleotides, particularly toxic genes, involves
PT   introducing a library of insert polynucleotides into a host cell
PT   population, where the target polynucleotide promotes cell death -
XX
PS   Example 1; Page 136; 359pp; English.
XX
CC   The present invention relates to a method for selecting a target
CC   polynucleotide. The method comprises introducing into a host cell
CC   population a library of insert polynucleotides, where expression of the
CC   target polynucleotide directly or indirectly promotes host cell death.
CC   The cells are cultured and the insert polynucleotides are collected from
CC   the cells which die. The method is useful for selecting target
CC   polynucleotides, particularly polynucleotides which alter cell phenotypes
CC   of induce or inhibit cell death. The method can be used to isolate toxic
CC   genes such as tumour suppressors. The present sequence was used to
CC   illustrate the method of the the present invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 29
AAE13119
ID   AAE13119 standard; peptide; 8 AA.
XX
AC   AAE13119;
XX
DT   28-JAN-2002  (first entry)
XX
DE   Ovalbumin (OVA)-derived immunodominant octapeptide.
XX
KW   Immunostimulatory fusion protein; IFP; antigen component; therapy;
KW   immunostimulatory component; T-cell mediated immune response; DC;
KW   dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
KW   Ovalbumin; OVA-derived immunodominant octapeptide.
XX
OS   Unidentified.
XX
PN   WO200174855-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US10515.
XX
PR   30-MAR-2000; 2000US-193504P.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Vidovic D,  Graddis T;
XX
DR   WPI; 2001-662965/76.
XX
PT   An immunostimulatory fusion protein comprising the intracellular domain
PT   of HER-2 and an antigen elicits an immune response to the antigen and
PT   is useful for the treatment of associated cancer associated -
XX
PS   Example 1; Page 25; 59pp; English.
XX
CC   The invention relates to immunostimulatory fusion proteins (IFP) and
CC   nucleic acid molecules encoding such proteins. The IFPs comprise a
CC   polypeptide antigen component and an immunostimulatory component derived
CC   from the intracellular domain of HER-2 protein which is effective to
CC   elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
CC   immune response to the antigen. IFP or superactivated dendritic cells
CC   are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
CC   associated with a particularly antigen. The present sequence is a
CC   ovalbumin (OVA)-derived immunodominant octapeptide. This peptide
CC   is used in the fusion constructs of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 30
AAE12145
ID   AAE12145 standard; peptide; 8 AA.
XX
AC   AAE12145;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Murine ovalbumin (OVA) peptide.
XX
KW   Microbial delivery vehicle; prophylactic; immunisation; gene therapy;
KW   tumour; carcinoma; neurodegeneration; muscular atrophy; cytostatic;
KW   neuroprotective; antibacterial; insecticide; fungicide; antiviral;
KW   antiprotozoal; cytostatic; anti-inflammatory; murine; ovalbumin; OVA;
KW   listeriolysin O; LLO; MHC; major histocompatibility complex.
XX
OS   Mus sp.
XX
PN   US6287556-B1.
XX
PD   11-SEP-2001.
```

```
XX
PF   21-DEC-1999;   99US-0469197.
XX
PR   13-AUG-1998;   98US-0133914.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Portnoy DA,  Higgins DE;
XX
DR   WPI; 2001-647179/74.
XX
PT   Vaccine for preventing (e.g. as immunizations) or treating (e.g. as
PT   gene therapy) tumor, carcinoma, neurodegeneration or muscular atrophy,
PT   comprises a non-virulent bacterium  -
XX
PS   Example; Column 12; 14pp; English.
XX
CC   The invention relates to microbial-based intracellular delivery of
CC   agents to eukaryotic cells. The agents include microbial delivery
CC   vehicles such as nonvirulent bacteria comprising a first gene
CC   encoding a nonsecreted foreign cytolysin operably linked to a
CC   heterologous promoter and a second gene encoding a different
CC   foreign agent. The foreign agent may be a nucleic acid or protein,
CC   and is frequently bioactive in and therapeutic to the target
CC   eukaryote. The vaccine comprising nonvirulent bacterium is useful
CC   for prophylactics (e.g. as immunisations) and treatments (e.g. as
CC   gene therapy) of e.g. tumour, carcinoma, neurodegeneration or
CC   muscular atrophy. The present sequence is murine ovalbumin (OVA)
CC   peptide. This sequence is used to examine the ability of
CC   Escherichia coli expressing listeriolysin O (LLO) and an antigenic
CC   protein to deliver the antigen to the cytosol of macrophages for
CC   processing and presentation on MHC (major histocompatibility
CC   complex) class I molecules.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 31
AAE09514
ID   AAE09514 standard; peptide; 8 AA.
XX
AC   AAE09514;
XX
DT   19-NOV-2001  (first entry)
XX
DE   Human ovalbumin peptide.
XX
KW   Mucin; cytostatic; immunostimulant; cell mediated immune response;
KW   carcinoma; adenocarcinoma; breast cancer; dendritic cell; vaccine;
KW   gene therapy; human; ovalbumin.
XX
OS   Homo sapiens.
XX
PN   WO200157068-A1.
XX
PD   09-AUG-2001.
XX
PF   01-FEB-2001; 2001WO-AU00090.
XX
PR   01-FEB-2000; 2000AU-0005369.
PR   14-JUN-2000; 2000US-0593870.
XX
PA   (AUST-) AUSTIN RES INST.
XX
PI   Mckenzie IFC,  Pietersz GA,  Apostolopoulous V;
XX
DR   WPI; 2001-541537/60.
XX
PT   Immunostimulant peptide, used as an anti-carcinoma vaccine, comprises a
PT   an epitope of the non-VNTR, non-leader region of a mucin -
XX
PS   Disclosure; Page 14; 84pp; English.
XX
CC   The patent discloses peptide or polypeptides capable of eliciting
CC   an immune response, comprising an amino acid sequence corresponding
CC   to an epitope of the non-central portion of varying numbers of an
CC   amino acid motif (VNTR), non-leader region of a mucin. The peptides
CC   of the invention, fusion proteins comprising the peptide and conjugate
CC   compounds with carbohydrate polymers are used to induce a cell mediated
CC   immune response against mucin in the prevention or treatment of
CC   carcinoma, preferably adenocarcinoma, most preferably breast cancer.
CC   They are also used to pulse dendritic cell for in vivo transfer and
CC   use as a vaccine. They are also used in gene therapy. The present
CC   sequence is ovalbumin peptide from human. This sequence is used for
CC   the prediction of T-cell epitopes.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 32
AAG63855
ID   AAG63855 standard; peptide; 8 AA.
XX
AC   AAG63855;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Amino acid sequence of an OVA-derived minimal CTL peptide.
XX
KW   OprI; lipoprotein; adjuvant; type1 immune response; gp63;
KW   Leishmania major; leishmaniasis; TBC; leprosy; mycotin infection;
KW   allergic asthma; autoimmune disease.
XX
OS   Synthetic.
XX
PN   WO200160404-A2.
XX
PD   23-AUG-2001.
XX
PF   13-FEB-2001; 2001WO-EP01673.
XX
PR   18-FEB-2000; 2000EP-0200589.
XX
PA   (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI   Revets H,  Cornelis P,  De Baetselier P;
XX
DR   WPI; 2001-522552/57.
XX
PT   Use of major OprI lipoprotein of Pseudomonas aeruginosa or its
PT   functional fragments as adjuvant to obtain a Th1 type immune response
PT   against heterologous antigen, for treating leishmaniasis, leprosy,
PT   allergic asthma  -
XX
PS   Disclosure; Page 12; 54pp; English.
XX
CC   The present sequence represents an OVA-derived minimal CTL peptide,
CC   which was used in the course of the invention. The specification
CC   describes the use of the major OprI lipoprotein of Pseudomonas
CC   aeruginosa or its functional fragments as an adjuvant to obtain
```

```
CC   a Th1 type immune response against a heterologous antigen. They are
CC   especially used as an adjuvant to obtain a Th1 type immune response
CC   against a heterologous antigen such as antigen gp63 of Leishmania major,
CC   for treating a disease such as leishmaniasis, TBC (undefined), leprosy,
CC   mycotin infection, allergic asthma or an autoimmune disease, in which
CC   the natural Th1 response is insufficient and/or in which the immune
CC   response is polarizes towards Th2 response.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 33
AAG66422
ID   AAG66422 standard; Peptide; 8 AA.
XX
AC   AAG66422;
XX
DT   23-OCT-2001  (first entry)
XX
DE   Chicken ovalbumin peptide, OVA257-264, used as a peptide antigen.
XX
KW   Immunomodulator; vaccine; immune response; immunogenic; chicken;
KW   ovalbumin.
XX
OS   Gallus domesticus.
XX
PN   WO200154720-A1.
XX
PD   02-AUG-2001.
XX
PF   05-JAN-2001; 2001WO-EP00087.
XX
PR   28-JAN-2000; 2000AT-0000129.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Lingnau K,  Mattner F,  Schmidt W,  Birnstiel M,  Buschle M;
XX
DR   WPI; 2001-536419/59.
XX
PT   Pharmaceutical composition useful for inducing immune response
PT   comprises antigen, immunogenic oligodeoxynucleotide containing
PT   cytosine-guanine dinucleotide motifs and polycationic polymer  -
XX
PS   Example 1; Page 22; 39pp; English.
XX
CC   The present invention relates to a pharmaceutical composition which
CC   comprises an antigen, an immunogenic oligodeoxynucleotide containing
CC   cytosine-guanine dinucleotide (CpG) motifs (CpG-ODN) and a polycationic
CC   polymer. The composition is useful for making a vaccine to induce potent
CC   immune responses, or to decrease or ablate undesired immune responses.
CC   The present sequence, OVA257-264, is a peptide from chicken ovalbumin.
CC   This sequence was used as a peptide antigen in the method of the present
CC   invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8
```

```
RESULT 34
AAU05398
ID   AAU05398 standard; peptide; 8 AA.
XX
AC   AAU05398;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Peptide released from ovalbumin (Ova) after cleavage of peptide P1.
XX
KW   Heat shock protein; hsp; CD8+ cytotoxic T lymphocyte; ovalbumin;
KW   CTL; CD4+ T cell; AIDS; acquired immunodeficiency syndrome; murine;
KW   human immunodeficiency virus; HIV; pathogen; cancer.
XX
OS   Mus sp.
XX
PN   WO200151081-A1.
XX
PD   19-JUL-2001.
XX
PF   01-DEC-2000; 2000WO-US32831.
XX
PR   14-JAN-2000; 2000US-0176143.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI   Huang Q,  Richmond JFL,  Cho BK,  Palliser D,  Chen J,  Eisen HN;
PI   Young RA;
XX
DR   WPI; 2001-451815/48.
XX
PT   Inducing a CD8+ cytotoxic T lymphocyte immune response in an individual
PT   for treating diseases such as HIV involves administering a fusion
PT   molecule comprising a heat shock protein -
XX
PS   Example 1; Fig 1C; 58pp; English.
XX
CC   The present sequence represents a naturally occurring murine peptide
CC   which is released from ovalbumin (Ova) upon cleavage of peptide P1
CC   (AAU05397). The present sequence is described in an invention
CC   relating to a novel method of inducing a CD8+ cytotoxic T lymphocyte
CC   (CTL) response to a molecule in an individual by administrating a
CC   fusion molecule joined to a hsp, or an adenosine triphosphate (ATP)
CC   binding domain of a hsp. The method is particularly useful in inducing
CC   a CD8+ CTL response in an individual deficient in CD4+ T cells e.g. for
CC   treating an AIDS acquired immunodeficiency syndrome patient carrying
CC   the human immunodeficiency virus (HIV). The method is also useful for
CC   treating diseases that are caused by or associated with intracellular
CC   pathogens, and for treating cancer.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 35
AAE06033
ID   AAE06033 standard; peptide; 8 AA.
XX
AC   AAE06033;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Chicken ovalbumin CTL epitope.
XX
```

```
KW   Chicken; vaccine; cytostatic; immunostimulant; antibacterial; antifungal;
KW   protozoacide; antiviral; recombinant flavivirus; immune response; TAA;
KW   tumour associated antigen; ovalbumin; epitope.
XX
OS   Gallus sp.
XX
PN   WO200153467-A1.
XX
PD   26-JUL-2001.
XX
PF   19-JAN-2001; 2001WO-US01866.
XX
PR   21-JAN-2000; 2000US-0177449.
PR   01-SEP-2000; 2000US-0653754.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Andino-pavlovsky R,  Mcallister-moreno A;
XX
DR   WPI; 2001-457605/49.
XX
PT   Recombinant flavivirus useful for reducing the number of tumor cells in
PT   a host comprises an exogenous nucleic acid encoding a polypeptide
PT   proteolytically processed after expression in the vector resulting in
PT   its release  -
XX
PS   Example 1; Page 19; 48pp; English.
XX
CC   The invention relates to a replication-competent recombinant flavivirus
CC   (yellow fever virus-YFV) comprising an exogenous nucleic acid encoding an
CC   exogenous polypeptide. The replication-competent recombinant flavivirus
CC   may be used to illicit an immune response to an antigen, e.g., tumour
CC   associated antigen (TAA) or a microbial pathogen antigen, in a mammalian
CC   host. Particularly, recombinant flavivirus is useful for reducing the
CC   number of tumour cells in a host. In addition, recombinant flavivirus
CC   may be useful as a vaccine to provide immune protection against pathogens
CC   such as bacteria, viruses, fungi and parasites. Unlike other vectors
CC   which will produce only one cycle of antigen expression and/or which
CC   will stop expression without the intervention of the host immune system,
CC   recombinant flaviviruses continue to propagate until the immune system
CC   is sufficiently activated to halt the infection. This produces a stronger
CC   immune response against the exogenous antigenic peptide produced from
CC   the flavivirus as compared to the immune response that would be elicited
CC   using conventional expression vectors (e.g., a viral replicon). The
CC   present peptide sequence is chicken ovalbumin CTL epitope which is the
CC   exogenous polypeptide.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 36
AAB84316
ID   AAB84316 standard; peptide; 8 AA.
XX
AC   AAB84316;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Peptide used to produce lemA peptide variants.
XX
XX
KW   lemA; CD8+ epitope; T cell response.
XX
OS   Synthetic.
XX
PN   WO200140275-A2.
XX
PD   07-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-US33027.
XX
PR   06-DEC-1999;   99US-0169227.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kurlander RJ,  Chao E,  Fields J;
XX
DR   WPI; 2001-389952/41.
XX
PT   New isolated variant of lemA, tlemA, comprising a hydrophobic element
PT   joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
PT   response or as a treatment or prophylactic against diseases  -
XX
PS   Disclosure; Page 6; 65pp; English.
XX
CC   The specification describes a peptide variant of lemA, comprising a
CC   hydrophobic element joined to a CD8+ epitope. The peptides may be
CC   used therapeutically by administering the peptides to a patient having
CC   a need to induce a directed CD8+ T cell response. The peptide may also
CC   be used as a preventive measure to avoid a disease or condition, or to
CC   treat subjects already afflicted with a disease. The present sequence
CC   can be inserted near the lemA hydrophobic element, to create peptides
CC   of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 37
AAB99354
ID   AAB99354 standard; Peptide; 8 AA.
XX
AC   AAB99354;
XX
DT   24-AUG-2001  (first entry)
XX
DE   Ovalbumin cytotoxic T lymphocyte epitope SEQ ID NO:11.
XX
KW   Immunostimulatory sequence; ISS; immunomodulatory; immune response;
KW   antigen; antiallergic; modulation; Th1 lymphocyte stimulation; allergy;
KW   Th1-associated cytokine; Th2 lymphocyte suppression; cytokine.
XX
OS   Synthetic.
XX
PN   WO200135991-A2.
XX
PD   25-MAY-2001.
XX
PF   15-NOV-2000; 2000WO-US31385.
XX
PR   15-NOV-1999;   99US-0165467.
PR   14-NOV-2000; 2000US-0713136.
XX
PA   (DYNA-) DYNAVAX TECHNOLOGIES CORP.
XX
PI   Tuck S,  Van Nest G;
XX
DR   WPI; 2001-329209/34.
XX
PT   Populations of conjugate molecules comprising polynucleotide
PT   immunostimulatory sequences polynucleotides and antigens, useful for
```

```
PT   controlling immune responses -
XX
PS   Example 6; Page 69; 97pp; English.
XX
CC   The present invention describes immunomodulatory populations ((I) and
CC   (II)) of conjugate molecules (CMs) comprising immunostimulatory sequences
CC   (ISS) of polynucleotides and antigens. The extent of conjugation affects
CC   the immunological properties (e.g. the extent of antigen-specific
CC   antibody formation, including Th1-associated antibody formation) so the
CC   conjugates are used for altering the type and extent of immune response.
CC   (I) and (II) have immunomodulatory, immunosuppressive and antiallergic
CC   activities, and can be used in the modulation of immune responses via
CC   the stimulation of Th1 lymphocytes and Th1-associated cytokines, and
CC   suppression of Th2 lymphocytes and cytokines. The populations ((I) and
CC   (II)) of conjugate molecules may be used for modulating immune responses
CC   in individuals e.g. for the treatment of an allergic condition. (I) and
CC   (II) may be used to modulate immune responses and therefore prevent
CC   potentially harmful reactions to antigens. The present sequence
CC   represents an ovalbumin (OVA) cytotoxic T lymphocyte (CTL) epitope
CC   which is used in the exemplification of the present invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 38
AAB82176
ID   AAB82176 standard; peptide; 8 AA.
XX
AC   AAB82176;
XX
DT   20-JUL-2001  (first entry)
XX
DE   Immunodominant CTL epitope of ovalbumin.
XX
KW   Vaccine; Antiviral; Antibacterial; Antiparasitic; liposome;
KW   archaeobacterium; cytotoxic T lymphocyte response;
KW   immunodominant epitope; ovalbumin; archaeosome.
XX
OS   Unidentified.
XX
PN   WO200126683-A2.
XX
PD   19-APR-2001.
XX
PF   12-OCT-2000; 2000WO-CA01197.
XX
PR   12-OCT-1999;   99US-0158944.
PR   08-JUN-2000; 2000US-0209988.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Sprott GD,  Krishnan L,  Conlan JW,  Omri A,  Patel GB;
XX
DR   WPI; 2001-281839/29.
XX
PT   New vaccine comprising a liposome useful for conferring protective
PT   immunity against an intracellular pathogen -
XX
PS   Disclosure; Page 34; 98pp; English.
XX
CC   The present invention relates to a vaccine composition comprising a
CC   liposome prepared from the total polar lipids extract of an
CC   archaebacterium and an acellular antigen, preferably an isolated
CC   outer membrane from a pathogen. The vaccine of the invention provides an
CC   enhanced cytotoxic T lymphocyte response. The vaccine of the invention
CC   is useful for conferring protective immunity against an intracellular
CC   pathogen. The present peptide: immunodominant CTL epitope of ovalbumin,
CC   was used to illustrate the present invention. This peptide was used to
CC   test for the ability of archaeosomes to induce CTL responses to
CC   ovalbumin.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 39
AAB81122
ID   AAB81122 standard; peptide; 8 AA.
XX
AC   AAB81122;
XX
DT   04-JUL-2001  (first entry)
XX
DE   Chicken ovalbumin (OVA) peptide.
XX
KW   Ovalbumin; chicken; OVA; immune response; Langerhans cell migration;
KW   tumour; EG7-OVA.
XX
OS   Gallus gallus.
XX
PN   US6210672-B1.
XX
PD   03-APR-2001.
XX
PF   20-OCT-1998;   98US-0176044.
XX
PR   20-OCT-1998;   98US-0176044.
XX
PA   (TORR-) TORREY PINES INST MOLECULAR STUDIES.
XX
PI   Cowing C;
XX
DR   WPI; 2001-280845/29.
XX
PT   Enhancing an immune response to an antigen in a mammal comprises
PT   topically administering the antigen, a penetration enhancer and an
PT   agent for enhancing Langerhans cell migration  -
XX
PS   Disclosure; Column 4; 15pp; English.
XX
CC   This invention relates to a method for enhancing an immune response to an
CC   antigen in a mammal. The method comprises administering a composition
CC   comprising the antigen, a penetration enhancer selected from lipophilic
CC   solvents, low-frequency ultrasound, electroporation, iontophoresis and
CC   intraepidermal delivery and an agent for enhancing Langerhans cell
CC   migration to an epidermal or mucosal site. The method can be used to
CC   enhance the immune response to tumours, viruses, bacteria and parasites.
CC   The present sequence represents a fragment of the chicken ovalbumin (OVA)
CC   protein. The peptide functions as a EG7-OVA tumour associated peptide
CC   antigen for CD8+ cytotoxic T lymphocytes. The peptide can be used in the
CC   method of the invention to enhance an immune response to the EG7-OVA
CC   tumour.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
```

```
Db          1 SIINFEKL 8


RESULT 40
AAB82065
ID   AAB82065 standard; peptide; 8 AA.
XX
AC   AAB82065;
XX
DT   22-JUN-2001  (first entry)
XX
DE   Ovalbumin-derived peptide, used as a control peptide.
XX
KW   Antigen; immunostimulant; vaccine; pharmaceutical composition; antiviral;
KW   viral infection; ovalbumin.
XX
OS   Unidentified.
XX
PN   WO200124822-A2.
XX
PD   12-APR-2001.
XX
PF   02-OCT-2000; 2000WO-EP09657.
XX
PR   01-OCT-1999;   99AT-0001680.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Fleitmann J,  Mattner F,  Buschle M,  Melling J;
XX
DR   WPI; 2001-290577/30.
XX
PT   New pharmaceutical composition comprising an antigen, an
PT   immunostimulating substance and a polycationic polymer, useful in
PT   manufacturing vaccines  -
XX
PS   Example 1; Page 11; 20pp; English.
XX
CC   The present invention relates to a pharmaceutical composition comprising
CC   (a) an antigen; (b) an immunostimulating substance consisting of
CC   neuroactive compounds, hormones, compounds having growth hormone activity
CC   or their mixtures; and (c) a polycationic polymer. The composition is
CC   useful in manufacturing vaccines. To illustrate the present invention, a
CC   murine tyrosinase related protein-2 peptide (TRP-2 peptide; see
CC   AAB82064), was used. Mice were injected subcutaneously with either the
CC   TRP-2 peptide, TRP-2 peptide + human growth hormone (HGH), TRP-2 peptide
CC   + poly-L-arginine 60 (pR60) or TRP-2 peptide + pR60 + HGH. Animals were
CC   sacrificed 10 days post injection, and mesenteric and inguinal lymph
CC   nodes were harvested. Lymphocytes were prepared from lymph nodes and were
CC   re-stimulated with TRP-2 peptide or with an ovalbumin-derived peptide
CC   (the present peptide), with the same major histocompatibility complex
CC   (MHC) restriction serving as negative control. Spots representing single
CC   T cells specific for the peptide used for re-stimulation were counted. No
CC   spots were detected when the ovalbumin derived peptide was used, while
CC   TRP-2 peptide + pR60 + HGH showed the highest number of spots or single T
CC   cells. The present peptide was also used as a control peptide for
CC   experiments with substance P (see AAB82070).
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 41
AAB92374
ID   AAB92374 standard; Peptide; 8 AA.
XX
AC   AAB92374;
XX
DT   22-JUN-2001  (first entry)
XX
DE   Miscellaneous peptide SEQ ID NO:1550.
XX
KW   Protection; endogenous therapeutic peptide; peptidase; conjugation;
KW   blood component; modification; succinimidyl; maleimido group; amino;
KW   hydroxyl; thiol; hormone; growth factor; neurotransmitter.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200069900-A2.
XX
PD   23-NOV-2000.
XX
PF   17-MAY-2000; 2000WO-US13576.
XX
PR   17-MAY-1999;   99US-0134406.
PR   10-SEP-1999;   99US-0153406.
PR   15-OCT-1999;   99US-0159783.
XX
PA   (CONJ-) CONJUCHEM INC.
XX
PI   Bridon DP,  Ezrin AM,  Milner PG,  Holmes DL,  Thibaudeau K;
XX
DR   WPI; 2001-112059/12.
XX
PT   Modifying and attaching therapeutic peptides to albumin prevents
PT   peptidase degradation, useful for increasing length of in vivo activity
PT   -
XX
PS   Disclosure; Page 711; 733pp; English.
XX
CC   The present invention describes a modified therapeutic peptide (I)
CC   comprising a therapeutically active amino acid region (III) and a
CC   reactive group (II) (e.g. succinimidyl and maleimido groups) attached to
CC   a less therapeutically active amino acid region (IV), which covalently
CC   bonds with amino/hydroxyl/thiol groups on blood components to form a
CC   peptidase stabilised therapeutic peptide composed of 3-50 amino acids.
CC   (I) are useful for modifying therapeutic peptides e.g. hormones, growth
CC   factors and neurotransmitters, to protect them from peptidase activity
CC   in vivo for the treatment of various disorders. Endogenous therapeutic
CC   peptides are not suitable as drug candidates as they require frequent
CC   administration due to rapid degradation by peptidases in the body.
CC   Modifying and attaching therapeutic peptides to albumin prevents or
CC   reduces the action of peptidases to increase length of activity (half
CC   life) and specificity as bonding to large molecules decreases
CC   intracellular uptake and interference with physiological processes.
CC   AAB90829 to AAB92441 represent peptides which can be used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 42
AAB48950
ID   AAB48950 standard; Protein; 8 AA.
XX
AC   AAB48950;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Ovalbumin MHC class I epitope, SEQ ID NO:6.
```

```
XX
KW   Transposable element; MHC epitope; major histocompatibility complex;
KW   intracellular bacterial pathogen; loxP site; Cre recombinase;
KW   insertion end; in-frame fusion; detection; antigen;
KW   disseminated insertions of class-I epitopes; DICE-I; transposon Tn5;
KW   ovalbumin MHC class I epitope.
XX
OS   Unidentified.
XX
PN   WO200071158-A1.
XX
PD   30-NOV-2000.
XX
PF   26-MAY-2000; 2000WO-US14687.
XX
PR   26-MAY-1999;   99US-0136210.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Heffron FL,  Parker DC,  Ellefson DD;
XX
DR   WPI; 2001-031967/04.
XX
PT   Transposable element for detecting an antigenic epitope of a pathogen,
PT   comprising 5' and 3' recombining sites, nucleic acid sequences encoding
PT   a selectable marker and major histocompatibility complex (MHC) epitope,
PT   and an insertion end -
XX
PS   Claim 5; Page 41; 63pp; English.
XX
CC   The invention relates to a novel transposable element comprising DNA
CC   encoding a selectable marker (e.g., antibiotic resistance) located
CC   between a 5' recombining site and a 3' recombining site (e.g., loxP
CC   sites); DNA encoding an MHC (major histocompatibility complex) epitope
CC   either 5' of the 5' recombining site or 3' of the 3' recombining site;
CC   and insertion ends comprising an inverted repeat sequence at the 5' and
CC   3' ends of the transposable element sufficient for integration of the
CC   transposable element. The transposable elements of the invention are able
CC   to introduce in-frame insertions throughout the chromosome of an
CC   intracellular bacterial pathogen. This system "tags" the bacterial gene
CC   and resulting protein, allowing the identification of proteins
CC   secreted across the membranes of the eukaryotic cell infected by the
CC   bacterium. In one embodiment, the transposable elements contain an
CC   antibiotic resistance cassette, two minimal loxP recombination sites, an
CC   MHC class I or class II epitope, and flanking insertion ends. A
CC   transposase, such as the Cre recombinase protein, is expressed in trans
CC   from a plasmid, or can be included in the transposable element. The Cre
CC   recombinase loops out the intervening sequences containing the antibiotic
CC   resistance cassette. When the transposable element inserts within a gene,
CC   the resolved insertion places the MHC class I or class II epitope in
CC   frame with the gene. The transposable elements of the invention are
CC   useful for detecting an antigenic epitope of an intracellular bacterial
CC   pathogen, such as Salmonella sp., Mycobacterium tuberculosis and Listeria
CC   monocytogenes. Certain embodiments of the technology, termed
CC   "disseminated insertions of class-I epitopes" (DICE-I; DICE-II for
CC   class II epitopes) allow the rapid and accurate identification of
CC   proteins involved in bacterial pathogenesis so that such proteins can
CC   be used as vaccine and drug targets. Carrier vaccines may be generated
CC   by infecting bacteria with a transposable element of the invention
CC   which additionally comprises an antigen associated with a disease,
CC   preferably cancer or a viral or bacterial disease, operably linked to the
CC   MHC epitope DNA of the transposable element. The present sequence
CC   represents an ovalbumin MHC class I epitope specifically claimed
CC   for use in the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8

RESULT 43
AAE28959
ID   AAE28959 standard; peptide; 8 AA.
XX
AC   AAE28959;
XX
DT   27-JAN-2003  (first entry)
XX
DE   Chicken ovalbumin peptide.
XX
KW   Modified vaccinia Ankara virus; MVA; HIV; human immunodeficiency virus;
KW   CD8+ T cell; immune response; acquired immune deficiency syndrome; AIDS;
KW   viral infection; vaccine; immunostimulant; virucide; chicken.
XX
OS   Gallus sp.
XX
PN   WO200272754-A2.
XX
PD   19-SEP-2002.
XX
PF   01-MAR-2002; 2002WO-US06713.
XX
PR   08-MAR-2001; 2001US-274434P.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Moss B,  Wyatt L,  Earl P;
XX
DR   WPI; 2002-723330/78.
XX
PT   New recombinant modified vaccinia Ankara (MVA) virus expressing HIV
PT   env, gag and pol genes, useful for boosting or inducing CD8 T cell
PT   immune responses in primates, e.g. humans, particularly for preventing
PT   AIDS or other viral infections -
XX
PS   Example 1; Page 111; 112pp; English.
XX
CC   The present invention relates to a composition comprising recombinant
CC   modified vaccinia Ankara (MVA) virus expressing an HIV (human immuno-
CC   deficiency virus) env, gag and pol gene or its modified gene for the
CC   production of an HIV Env, Gag and Pol antigen by expression from the
CC   recombinant MVA virus. The HIV env gene is modified to encode an HIV
CC   Env protein composed of gp120 and the membrane-spanning and ectodomain
CC   of gp41, but lacking part or all of the cytoplasmic domain of gp41. The
CC   composition or recombinant MVA virus is useful for boosting or inducing
CC   CD8+ T cell immune response in primates, particularly in humans. The
CC   composition may be used for preventing AIDS (acquired immune deficiency
CC   syndrome) or other viral infections. Sequences of the invention are also
CC   used as vaccines. The present sequence is chicken ovalbumin peptide. This
CC   sequence is used in the exemplification of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8

RESULT 44
AAE26368
ID   AAE26368 standard; peptide; 8 AA.
XX
AC   AAE26368;
XX
DT   13-DEC-2002  (first entry)
XX
```

```
DE   Ovalbumin CTL epitope.
XX
KW   Human; immune response; T-helper cell epitope; chitosan; CTL response;
KW   vaccine; prostate cancer; breast cancer; cytostatic; immunostimulant.
XX
OS   Unidentified.
XX
PN   WO200234287-A2.
XX
PD   02-MAY-2002.
XX
PF   26-OCT-2001; 2001WO-DK00705.
XX
PR   27-OCT-2000; 2000DK-0001606.
PR   03-NOV-2000; 2000US-245166P.
PR   18-JUN-2001; 2001DK-0000936.
XX
PA   (PHAR-) PHARMEXA AS.
XX
PI   Beier AM,  Gautam A,  Mouritsen S;
XX
DR   WPI; 2002-463339/49.
XX
PT   Inducing or enhancing an immune response against an antigen,
PT   particularly cytotoxic T-lymphocyte responses, for treating or
PT   ameliorating prostate or breast cancer, comprises administering the
PT   antigen formulated with chitosan  -
XX
PS   Example 3; Page 63; 97pp; English.
XX
CC   The invention relates to a method for inducing or enhancing an immune
CC   response against a polypeptide antigen in an animal, including human.
CC   The method comprises administering the polypeptide antigen or at least
CC   one variant which includes at least one first T-helper cell epitope that
CC   is foreign to the animal (foreign TH epitope) and is formulated with
CC   chitosan. The polypeptide antigen is weakly immunogenic or non-
CC   immunogenic. The invention is used as vaccine. The chitosan and
CC   polypeptide antigen or its variant are useful in the preparation of an
CC   immunogenic composition for inducing or enhancing an immune response,
CC   particularly CTL response, against the polypeptide or protein antigen.
CC   The method for inducing or enhancing an immune response is useful in
CC   treating or ameliorating cancer, e.g. prostate or breast cancer. The
CC   present sequence is ovalbumin CTL epitope used to illustrate the method
CC   of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8

RESULT 45
ABB79933
ID   ABB79933 standard; Peptide; 8 AA.
XX
AC   ABB79933;
XX
DT   12-DEC-2002  (first entry)
XX
DE   Ovalbumin T-cell epitope.
XX
KW   Vaccine; genetic immunisation; gene therapy; antigen; epitope;
KW   T-cell; T-lymphocyte; ovalbumin.
XX
OS   Unidentified.
XX
PN   US2002115625-A1.
XX
PD   22-AUG-2002.
XX
PF   08-MAR-2001; 2001US-0801540.
XX
PR   19-MAY-1999;   99US-0308511.
XX
PA   (BOTA/) BOT A.
PA   (BONA/) BONA C.
XX
PI   Bot A,  Bona C;
XX
DR   WPI; 2002-712482/77.
XX
PT   Immunizing an infant mammal against a target antigen or inducing a
PT   cytotoxic T cell response against a pathogen in the mammal, comprises
PT   administering nucleic acid encoding relevant epitopes of pathogen
PT   associated target antigens  -
XX
PS   Disclosure; Page 4; 45pp; English.
XX
CC   The present sequence is that of an ovalbumin T-cell epitope (amino
CC   acids 257-264).  This is a T-CTL epitope which, in the context of
CC   MHC Class II self antigens, may be recognised by a cytotoxic T-cell
CC   and thereby promote CTL-mediated lysis of cells comprising the target
CC   antigen.  It is an example of T-cell epitopes which may be used
CC   according to the invention.  The invention relates to immunising an
CC   infant mammal against a target antigen or inducing a cytotoxic
CC   T-cell response against a pathogen.  The method involves inoculating
CC   the infant with a nucleic acid encoding one or more relevant epitopes
CC   of one or more target antigens associated with the pathogen in a
CC   carrier, so that the relevant epitope(s) is expressed in the infant
CC   mammal.  B- or T-cell epitopes may be used, and the pathogen may be
CC   a virus, bacterium, protozoan, fungus, yeast, or parasite.  The method
CC   may reduce the need for subsequent boost administrations and may
CC   prevent the side-effects associated with live attenuated vaccines.
CC   Administration of multiple epitopes directed to antigens
CC   associated with more than one pathogen may provide an infant with a
CC   broader spectrum of protection, and may be a means for inducing an
CC   immune response to a variety of childhood pathogens.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8

RESULT 46
ABG93028
ID   ABG93028 standard; Peptide; 8 AA.
XX
AC   ABG93028;
XX
DT   20-NOV-2002  (first entry)
XX
DE   Mouse class 1 MHC molecule Kb binding ovalbumin epitope.
XX
KW   Regulator; transcription; cell death; phenotype; molecular scaffold;
KW   gene therapy; cancer; cardiovascular disease; arrhythmia; heart failure;
KW   ischaemia; obesity; neurodegenerative disease; Alzheimer's disease;
KW   bone pathology; dermatologic disease; psoriasis; infection; AIDS;
KW   acquired immunodeficiency syndrome; cosmetic; wound healing;
KW   antibiotic transport; drug toxicity; drug resistance; immunobiology;
KW   inflammation; allergic response; human immunodeficiency virus.
XX
OS   Unidentified.
XX
PN   WO200262822-A2.
XX
```

```
PD   15-AUG-2002.
XX
PF   04-FEB-2002; 2002WO-US02814.
XX
PR   02-FEB-2001; 2001US-265589P.
PR   05-FEB-2001; 2001US-265880P.
PR   27-FEB-2001; 2001US-271423P.
PR   23-JAN-2001; 2001US-263226P.
PR   28-MAR-2000; 2000US-192586P.
PR   22-SEP-1997;   97US-935377P.
XX
PA   (UYRP ) UNIV ROCHESTER.
XX
PI   Zauderer M,  Smith ES;
XX
DR   WPI; 2002-643398/69.
XX
PT   Identifying regulator polypeptides which influence target
PT   transcriptional regulatory regions, useful for treating cancer,
PT   comprises introducing host cells expressing the polypeptide into a
PT   library of polynucleotides  -
XX
PS   Example 1; Page 82; 224pp; English.
XX
CC   The invention discloses a method for identifying polynucleotides encoding
CC   a regulator polypeptide, whose expression induces activation of a target
CC   transcriptional regulatory region in a host cell. The method comprises
CC   providing a population of eukaryotic host cells capable of expressing the
CC   polypeptide, introducing into the host cell a library of polynucleotides
CC   encoding the polypeptides, permitting expression of the polypeptides and
CC   then recovering them from the host cells. The target transcriptional
CC   regulatory region is operably associated with a polynucleotide encoding a
CC   gene product, the expression of which results in host cell death or cause
CC   the host cells to exhibit a pre-determined modified phenotype and where
CC   the gene product is expressed upon activation of target transcriptional
CC   regulatory region. Each candidate regulator polypeptide comprises a
CC   candidate peptide and a molecular scaffold fused to the peptide so that
CC   the peptide is displayed on the surface of the candidate regulator
CC   polypeptide. The methods are useful in selecting and/or screening
CC   regulator molecules, such as polypeptides, which directly or indirectly
CC   induce or suppress the transcriptional activation of a target
CC   transcriptional regulatory region in a eukaryotic host cell. These
CC   regulator molecules may be used (e.g. in gene therapy) for preventing or
CC   treating cancers (e.g. breast or ovarian cancer), cardiovascular diseases
CC   (e.g. arrhythmia, heart failure, ischaemia), obesity, neurodegenerative
CC   diseases (e.g. Alzheimer's disease), bone pathologies, dermatologic
CC   diseases (e.g. psoriasis), infections (e.g viral, bacterial), acquired
CC   immunodeficiency syndrome (AIDS), in cosmetic applications and in wound
CC   healing. The method is also useful in screening regulator molecules that
CC   block antibiotic transport mechanisms, in drug toxicities and drug
CC   resistance applications and in improving the performance of existing or
CC   developmental drugs. It may also be used in immunobiology, inflammation,
CC   allergic response and in biotechnology applications. The sequences
CC   presented in ABG92946-ABG93029 are examples of regulator polypeptides.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 47
AAE25400
ID   AAE25400 standard; peptide; 8 AA.
XX
AC   AAE25400;
XX
DT   30-OCT-2002  (first entry)
XX
DE   Ovalbumin peptide used in the invention.
XX
KW   Recombinant vector; coat protein; CP; viral replication; infection;
KW   Zucchini yellow mosaic potyvirus; ZYMV; cucurbit fruit; vaccination;
KW   pharmaceutical; diagnostic; ovalbumin.
XX
OS   Unidentified.
XX
PN   WO200244323-A2.
XX
PD   06-JUN-2002.
XX
PF   28-NOV-2001; 2001WO-IL01098.
XX
PR   28-NOV-2000; 2000US-253136P.
PR   27-SEP-2001; 2001US-0963761.
XX
PA   (VIRO-) VIROGENE LTD.
XX
PI   Gal-On A,  Shiboleth YM,  Arazi T,  Ilan Y;
XX
DR   WPI; 2002-537446/57.
DR   N-PSDB; AAD41429.
XX
PT   Novel recombinant vector useful for transiently expressing heterologous
PT   peptide in plant comprises potyvirus nucleic acid sequence and
PT   heterologous sequence inserted at amino terminus of potyvirus coat
PT   protein -
XX
PS   Claim 17; Page 60; 61pp; English.
XX
CC   The invention relates to a recombinant vector for expressing a
CC   heterologous peptide at the amino-terminus of a potyvirus coat protein
CC   (CP). The vector includes sufficient potyvirus nucleic acid sequence
CC   to permit viral replication and spread within a plant infected by the
CC   vector. The invention also relates to Zucchini yellow mosaic potyvirus
CC   (ZYMV) AGII strain CP and its corresponding nucleic acid sequence. The
CC   recombinant vector is useful for transiently expressing a portion of
CC   the heterologous peptide in a plant. It is also useful for infecting a
CC   cucurbit fruit, is useful as a source of material for vaccination,
CC   pharmaceutical or diagnostic application. The present sequence is a
CC   ovalbumin peptide used to fuse to the N-terminus of ZYMV AGII
CC   strain CP.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 48
ABG31661
ID   ABG31661 standard; Peptide; 8 AA.
XX
AC   ABG31661;
XX
DT   05-NOV-2002  (first entry)
XX
DE   Chicken ovalbumin MHC class I restricted epitope.
XX
KW   Chicken; ovalbumin; MHC class I; major histocompatibility complex;
KW   polycationic compound; allergen; cytokine; chemokine; wound healing;
KW   cytotoxic drug; anti-oligogenic drug; immunostimulant; antiallergic;
KW   cytostatic; vulnerary; immunogenic.
XX
OS   Gallus gallus.
XX
```

```
PN   WO200253184-A2.
XX
PD   11-JUL-2002.
XX
PF   07-JAN-2002; 2002WO-EP00062.
XX
PR   05-JAN-2001; 2001WO-EP00087.
PR   25-APR-2001; 2001AT-0000672.
XX
PA   (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Lingnau K,  Mattner F,  Schmidt W,  Birnstiel M,  Buschle M;
XX
DR   WPI; 2002-627324/67.
XX
PT   Use of a polycationic compound for the preparation of a medicament with
PT   retarded in vivo release  -
XX
PS   Example 1; Page 13; 29pp; English.
XX
CC   The invention relates to preparation of a medicament with retarded in
CC   vivo release comprising use of a polycationic compound. The compound is
CC   used in the preparation of a medicament with retarded in vivo release,
CC   and a vaccine containing an antigen. The medicament includes e.g. an
CC   allergen, a cytokine, a chemokine, an immunostimulatory nucleic acid, a
CC   cytotoxic or an anti-oligogenic drug or a compound needed for wound
CC   healing. The medicament prevents or ameliorates side effects of drugs,
CC   which are due to its too fast distribution of the drug throughout the
CC   body by exhibiting a retarded release of the drug from the site of
CC   administration. In the case of vaccine the compounds provide a depot,
CC   which allows a long lasting continuous and effective presentation of the
CC   antigen to the immune system to create a protective immunity. This
CC   sequence represents a chicken ovalbumin MHC class I restricted epitope
CC   used in the scope of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 49
ABG31967
ID   ABG31967 standard; Peptide; 8 AA.
XX
AC   ABG31967;
XX
DT   05-NOV-2002  (first entry)
XX
DE   Chicken ovalbumin OVA257-264-peptide.
XX
KW   Chicken; ovalbumin; polycationic; antiinflammatory; immunostimulant;
KW   antiallergic; cytostatic; vulnerary; medicament; inflammatory potential;
KW   inflammation; vaccine; antigen; adjuvant; allergen; cytokine; chemokine;
KW   immunostimulatory nucleic acid; cytotoxic drug; antioligogenic drug;
KW   wound healing; OVA-peptide; epitope; major histocompatibility complex;
KW   MHC.
XX
OS   Gallus gallus.
XX
PN   WO200253185-A2.
XX
PD   11-JUL-2002.
XX
PF   07-JAN-2002; 2002WO-EP00071.
XX
PR   05-JAN-2001; 2001WO-EP00087.
PR   25-APR-2001; 2001AT-0000670.
XX
PA   (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Lingnau K,  Egyed A,  Schmidt W,  Buschle M,  Grill S;
XX
DR   WPI; 2002-627325/67.
XX
PT   Use of a polycationic compound for the preparation of a medicament with
PT   reduced inflammatory potential  -
XX
PS   Example 1; Page 13; 43pp; English.
XX
CC   The invention discloses the use of a polycationic compound for the
CC   preparation of a medicament with reduced inflammatory potential, for
CC   treating or preventing inflammation or for a vaccine containing an
CC   antigen, possible acting as an adjuvant. The medicaments include
CC   allergens, cytokines, chemokines, immunostimulatory nucleic acids,
CC   cytotoxic or antioligogenic drugs and compounds needed for wound healing.
CC   The medicament acts locally at the site of administration, and lowers or
CC   completely eliminates inflammatory side effects of medicaments. Thus the
CC   medicament reduces the inflammatory potential of a medicament and allows
CC   the administration of medicaments that are usually not administered or
CC   only rarely administered due to their inflammatory side effects. The
CC   sequence presented is the OVA 257-264-peptide, a major histocompatibility
CC   complex (MHC) class I (H-2Kb) restricted epitope of chicken ovalbumin
CC   which was used in the scope of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 50
AAU99718
ID   AAU99718 standard; Peptide; 8 AA.
XX
AC   AAU99718;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Mouse MHC classI Kb OVA peptide sequence.
XX
KW   Mutant major histocompatibility complex class I chimeric protein; MHC;
KW   lymphocyte; T-cell receptor; tissue sample; biopsy material; pathogen;
KW   bodily fluid; T lymphocyte; neoplastic cell; tumour cell; MHC antigen;
KW   virus; protozoan; bacteria; fungi; nematode; immune response; activator;
KW   enhancer; T cell activator; mouse; recombinant yeast cell; Kb; OVA; Ld;
KW   beta2m; dEV8.
XX
OS   Mus sp.
XX
PN   WO200246399-A2.
XX
PD   13-JUN-2002.
XX
PF   10-DEC-2001; 2001WO-US47817.
XX
PR   08-DEC-2000; 2000US-254495P.
XX
PA   (UNII ) UNIV ILLINOIS FOUND.
XX
PI   Kranz DM,  Brophy S;
XX
DR   WPI; 2002-527916/56.
XX
```

```
PT   New isolated mutant major histocompatibility complex class I chimeric
PT   protein displayed on surfaces of recombinant yeast cells, has improved
PT   stability, and is useful for activating immune response  -
XX
PS   Example 3; Figure 18; 96pp; English.
XX
CC   The present invention relates to a new mutant major histocompatibility
CC   complex (MHC) class I chimeric protein. The protein of the invention
CC   comprises a portion mediating binding to surfaces of recombinant yeast
CC   cells and a portion comprising peptide binding region of MHC class I
CC   protein, where the invention is improved in stability as compared with
CC   MHC class I chimeric protein which is not a mutant chimeric protein.
CC   The protein, further comprising a detectable label, is useful for
CC   detecting a lymphocyte having a T-cell receptor protein in a biological
CC   sample such as cells, tissue sample, biopsy material or bodily fluids.
CC   The method is useful for detecting a T lymphocyte that is specific for
CC   a neoplastic cell, a tumour cell, a virus-infected cell, a protozoan-
CC   infected cell, a bacterium-infected cell or a fungus-infected cell. The
CC   protein of the invention can be used to directly activate T cells, in
CC   order to identify/screen for peptide-MHC antigens. The protein is also
CC   useful in activating T cells that participate in the removal of target
CC   cells including neoplastic cells and cells infected with pathogenic
CC   agents including viruses, protozoans, bacteria, fungi or nematodes.
CC   The invention is improved in stability as compared with MHC class I
CC   protein which is not a mutant chimeric protein. The present amino acid
CC   sequence represents a mouse MHC peptide of the invention, as described
CC   above.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                ||||||||
Db          1 SIINFEKL 8


RESULT 51
ABB08108
ID   ABB08108 standard; peptide; 8 AA.
XX
AC   ABB08108;
XX
DT   10-SEP-2002  (first entry)
XX
DE   Chicken ovalbumin CTL epitope fragment.
XX
KW   T cell; antigen; tumour; vaccine; cytostatic; cancer; ovalbumin;
KW   CTL epitope.
XX
OS   Gallus sp.
XX
PN   US6387701-B1.
XX
PD   14-MAY-2002.
XX
PF   30-APR-1999;   99US-0302329.
XX
PR   30-APR-1996;   96US-0640444.
PR   30-APR-1997;   97WO-US07317.
PR   06-MAY-1998;   98US-0073819.
PR   16-FEB-1999;   99US-0171916.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Nair SK,  Boczkowski DJ,  Gilboa E;
XX
DR   WPI; 2002-478447/51.
XX
PT   Identifying tumor antigens that elicit T cell responses and which may
PT   be used for vaccinating against cancers, e.g. melanomas, breast
PT   cancers, prostate cancers, colon cancers, and ovarian cancers,
PT   comprises a cytotoxicity assay  -
XX
PS   Example 1; Column 12; 21pp; English.
XX
CC   The invention relates to identifying a tumour antigen that elicits a T
CC   cell response directed against the tumour. The antigen may then be used
CC   to vaccinate against cancers e.g. melanomas, bladder cancers, breast
CC   cancers, pancreatic cancers. The present sequence represents a chicken
CC   ovalbumin CTL epitope fragment.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                ||||||||
Db          1 SIINFEKL 8


RESULT 52
ABB81273
ID   ABB81273 standard; peptide; 8 AA.
XX
AC   ABB81273;
XX
DT   20-AUG-2002  (first entry)
XX
DE   Chicken OVA 257-264 peptide SEQ ID NO:1.
XX
KW   Yeast; dendritic cell; vaccine; immune response; ovalbumin; antifungal;
KW   immunostimulant; antibacterial; virucide; antiprotozoal; cytostatic;
KW   immunisation; cell mediated immunity; infectious disease; cancer.
XX
OS   Gallus gallus.
XX
PN   WO200239951-A2.
XX
PD   23-MAY-2002.
XX
PF   15-NOV-2001; 2001WO-US43537.
XX
PR   15-NOV-2000; 2000US-249173P.
XX
PA   (GLOB-) GLOBE IMMUNE INC.
PA   (UYTE-) UNIV TECHNOLOGY CORP.
XX
PI   Duke RC,  Bellgrau D,  Franzusoff A,  Wilson CC;
XX
DR   WPI; 2002-479895/51.
XX
PT   Therapeutic composition, useful as vaccine, comprises dendritic cell
PT   intracellularly loaded with yeast vehicle and at least one antigen  -
XX
PS   Example 4; Page 34; 68pp; English.
XX
CC   The present invention describes a therapeutic composition (I) comprising
CC   a dendritic cell (II), a yeast vehicle (III) and at least one antigen
CC   (IV), where (II) has been loaded intracellularly with (III) and (IV).
CC   (I) has immunostimulant, antibacterial, antifungal, virucide,
CC   antiprotozoal and cytostatic activities. (I) has many attributes that
CC   make it an ideal vaccine candidate, including ease of construction, low
CC   expense of mass production, biological stability and safety. No grossly
CC   adverse side effects of immunisation with whole yeast were apparent at
CC   the time of the initial vaccination or upon real administration. The
CC   composition provides a powerful strategy for the induction of cell-
CC   mediated immunity directed against a variety of infectious diseases and
CC   cancer targets. The present sequence represents a chicken OVA (ovalbumin)
CC   257-264 peptide which is used in an example from the present invention.
XX
SQ   Sequence   8 AA;
```

```
  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 53
AAE22531
ID   AAE22531 standard; peptide; 8 AA.
XX
AC   AAE22531;
XX
DT   26-JUL-2002  (first entry)
XX
DE   Ovalbumin peptide to evaluate the ability of aroC/sifA (P3H8) TML mutant.
XX
KW   SifA; attenuated microorganism; medicament; allergen; gene therapy;
KW   vaccine; protein therapy; virucide; hepatotropic; antiinflammatory;
KW   protozoacide; ovalbumin.
XX
OS   Unidentified.
XX
PN   WO200226251-A1.
XX
PD   04-APR-2002.
XX
PF   01-OCT-2001; 2001WO-GB04358.
XX
PR   29-SEP-2000; 2000GB-0023906.
PR   14-AUG-2001; 2001GB-0019802.
XX
PA   (MICR-) MICROSCIENCE LTD.
XX
PI   Brennan FR,  Dougan G;
XX
DR   WPI; 2002-339986/37.
XX
PT   An attenuated Salmonella strain, for producing an elevated immune
PT   response treat diseases, comprises a mutation that disrupts expression
PT   of the sifA gene, and expresses a therapeutic heterologous peptide such
PT   as an antigen -
XX
PS   Example 1; Page 7; 24pp; English.
XX
CC   The present invention relates to an attenuated Salmonella microorganism
CC   which comprises a mutation that disrupts the expression of the sifA gene
CC   and expresses a therapeutic heterologous peptide such as an antigen. The
CC   attenuated microorganisms are useful for manufacturing medicaments to
CC   treat or prevent a disease which can be treated by the heterologous
CC   product and to increase the MHC class I-restricted response in a patient,
CC   to deliver a therapeutic polynucleotide to a host cell to treat a disease
CC   which can be corrected by administering the polynucleotide and to cause
CC   an increase in the MHC class I-restricted response in a patient. They
CC   are also useful to deliver heterologous antigens and allergens to a
CC   patient, such as hepatitis, herpes simplex and Malarial antigens. The
CC   method of the invention is useful for delivery of antisense nucleotides
CC   or ribozymes in gene therapy. Sequences of the invention are also used
CC   as vaccines and in protein therapy. The present sequence is ovalbumin
CC   peptide (OVA 257-264) used to evaluate the ability of an aroC/sifA (P3H8)
CC   TML mutant to stimulate a MHC class I-restricted response to heterologous
CC   antigen.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8

RESULT 54
ABB76050
ID   ABB76050 standard; Peptide; 8 AA.
XX
AC   ABB76050;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Ovalbumin, H-2kb restricted epitope.
XX
KW   Bordetella pertussis; adenylate cyclase; CyaA; adenylcyclase;
KW   vector; drug delivery; antigen delivery; cell targetting; CD11b;
KW   ovalbumin; epitope; chicken.
XX
OS   Gallus sp.
XX
PN   EP1188446-A1.
XX
PD   20-MAR-2002.
XX
PF   15-SEP-2000; 2000EP-0402562.
XX
PR   15-SEP-2000; 2000EP-0402562.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CENT NAT RECH SCI.
XX
PI   Leclerc C,  Guermonprez P,  Ladant D,  Guiso N,  Khelef N;
XX
DR   WPI; 2002-354020/39.
XX
PT   Use of Bordetella adenylcyclase to make proteinaceous vector, useful
PT   for drug or antigen delivery, selectively targets cells that express
PT   CD11b  -
XX
PS   Example B; Page 14; 34pp; English.
XX
CC   The present sequence is the peptide sequence of a chicken ovalbumin,
CC   H-2Kb restricted epitope, which was used as an experimental model
CC   epitope in an example from the invention.  The epitope was
CC   genetically inserted into the catalytic domain of a detoxified, but
CC   still invasive, mutant adenylate cyclase (adenylcyclase, CyaA) of
CC   Bordetella pertussis.  The recombinant toxin, CyaAOVA, was used to
CC   immunise C57BL/6(H-2b) mice once i.v.  CD4- and CD40-independent
CC   cytotoxic T lymphocyte (CTL) priming was observed in the absence
CC   of adjuvant.  The invention relates to the novel use of Bordetella
CC   CyaA as a proteinaceous vector for targetting a molecule of interest
CC   to the surface CD11b-expressing cells, especially dendritic cells
CC   and neutrophils.  The molecule of interest is translocated in the
CC   cytosol to prime a CTL response.  In a preferred embodiment, a
CC   peptide is inserted into in the catalytic domain of CyaA at a
CC   permissive site.  The peptide may be an intracellular bacterial
CC   cell, tumour, viral, fungal or parasite cell antigen (all claimed).
CC   Alternatively, a drug, especially an antiinflammatory, is chemically
CC   coupled to CyaA for drug delivery.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 55
AAE19945
```

```
ID   AAE19945 standard; peptide; 8 AA.
XX
AC   AAE19945;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Cytotoxic T-cell epitope for ovalbumin.
XX
KW   Cytotoxic T cell; CTL; tumour; cancer; infection; cell-mediated immunity;
KW   vaccine; immune response; cytostatic; T-cell epitope; ovalbumin.
XX
OS   Unidentified.
XX
PN   US2002018785-A1.
XX
PD   14-FEB-2002.
XX
PF   02-APR-2001; 2001US-0822250.
XX
PR   22-SEP-1997;   97US-0935377.
XX
PA   (UYRP ) UNIV ROCHESTER.
XX
PI   Zauderer M;
XX
DR   WPI; 2002-239252/29.
XX
PT   Representational Difference Analysis method for identification of
PT   antigens recognized by cytotoxic T cells and specific for human tumors,
PT   comprises improved selection of genes encoding target antigens  -
XX
PS   Example 1; Page 13; 54pp; English.
XX
CC   The present invention relates to novel methods for the identification
CC   of antigens recognised by cytotoxic T cells (CTLs) and specific for
CC   human tumours, cancers and infected cells. The method involves screening
CC   the products of an expression library generated from DNA/RNA of a cell
CC   expressing a target epitope with cytotoxic T cells generated against
CC   the cell to identify DNA clones expressing target epitope or providing
CC   cytotoxic T cells specific for a gene product differentially expressed
CC   by a cell and measuring the cross-reactivity of the cytotoxic T cells
CC   for cells expressing a target epitope in which the target epitope is
CC   identified as a gene product inducing cytotoxic T cells. The method is
CC   useful for identifying a target epitope or antigen specific for a tumour
CC   cell. The target epitope is also useful for identifying target antigens
CC   in other target cells against which it is desirable to induce cell-
CC   mediated immunity. The antigen identified by the method is useful
CC   in immunogenic compositions and vaccine preparations to induce the
CC   regression of tumours, cancers and infections in mammals. The invention
CC   also relates to vaccinia viral vectors which are useful for treating
CC   tumour-bearing mammals, including humans to generate immune response
CC   against the tumour cells. They are also useful for immunising or
CC   vaccinating tumour-free subjects to prevent tumour formation. The
CC   present sequence is cytotoxic T-cell epitope for ovalbumin. This
CC   peptide is used in the exemplification of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 56
ABB09907
ID   ABB09907 standard; peptide; 8 AA.
XX
AC   ABB09907;
XX
DT   10-JUN-2002  (first entry)
XX
DE   Ovalbumin peptide (H-2kb CTL epitope).
XX
KW   Ovalbumin; cytokine; ELISPOT assay; polycationic substance;
KW   cytokine secreting cell; interleukin; IL; interferon; IFN; TNF; CSF;
KW   tumour necrosis factor; colony stimulating factor;
KW   enzyme-linked imunosorbent spot assay.
XX
OS   Aves.
XX
PN   WO200179854-A1.
XX
PD   25-OCT-2001.
XX
PF   12-APR-2001; 2001WO-EP04208.
XX
PR   13-APR-2000; 2000AT-0000645.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Schalich J,  Buschle M;
XX
DR   WPI; 2002-240136/29.
XX
PT   Identification of cytokine secreting cells involves incubating cell
PT   suspension containing antigen specific cells in the presence of
PT   polycationic substance  -
XX
PS   Example 1; Page 10; 22pp; English.
XX
CC   The sequence represents an ovalbumin peptide (H-2kb CTL epitope from
CC   ovalbumin). The invention relates to a novel method for identifying
CC   cytokine secreting cells (especially ELISPOT assays) using a polycationic
CC   substance. The method is useful for identifying cytokine secreting cells.
CC   The cytokine may be selected from interleukins (IL), interferons (IFN),
CC   tumour necrosis factors (TNF), and colony stimulating factors (CSF).
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 57
AAU76942
ID   AAU76942 standard; Peptide; 8 AA.
XX
AC   AAU76942;
XX
DT   05-JUN-2002  (first entry)
XX
DE   OVA peptide (257-264).
XX
KW   ODN; immunostimulatory oligodeoxynucleotide; Ovalbumin; OVA;
KW   cardiant; vaccine; tuberculosis; diphtheria; pertussis; measles;
KW   tetanus; acquired immune deficiency syndrome; AIDS; malaria;
KW   cardiovascular disease; cancer; deoxyinosine; chicken.
XX
OS   Gallus sp.
XX
PN   WO200193905-A1.
XX
PD   13-DEC-2001.
XX
PF   07-JUN-2001; 2001WO-EP06433.
XX
PR   08-JUN-2000; 2000AT-0001000.
```

```
PR   23-NOV-2000; 2000AT-0001973.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Schmidt W,  Lingnau K,  Schellack C,  Egyed A;
XX
DR   WPI; 2002-240927/29.
XX
PT   New oligodeoxynucleic acid molecule useful as an immunostimulatory
PT   agent  -
XX
PS   Example 1; Page 18; 52pp; English.
XX
CC   This invention relates to immunostimulatory oligodeoxynucleic acid
CC   molecules (ODN) that can be used to enhance an immune response for
CC   use in vaccines. The immunostimulatory oligonucleotides of the invention
CC   and pharmaceutical compounds containing them may be used as medicine,
CC   especially as an immunostimulatory agent, for the preparation of
CC   vaccines useful for the treatment of tuberculosis, diphtheria,
CC   pertussis, measles and tetanus, acquired immune deficiency syndrome
CC   (AIDS), malaria, cardiovascular diseases, and cancer. Oligonucleotides
CC   containing deoxyinosine residues (I-ODN) show a better immunostimulatory
CC   effect compared to prior art compounds containing CpG motifs. The ODNs
CC   of the invention produce more specific immune response to a given
CC   antigen or antigen fragment than the prior art compounds containing CpG.
CC   Using immunostimulatory oligonucleotides containing deoxyinosine reduces
CC   the induction of adverse side reactions, especially the induction of
CC   systemic TNF-alpha or interleukin-6. The immunostimulatory effect of the
CC   composition containing a polycationic polymer and an antigenic fragment
CC   was significantly higher than could be expected from the addition of the
CC   effects of each single component or even the addition of the effects of
CC   the ODN or the polycation with the antigen. The present sequence
CC   represents the ovalbumin (OVA) peptide used as an antigen in
CC   examples of the method of the invention.
XX
SQ   Sequence   8 AA;

  Query Match            100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 58
AAU76802
ID   AAU76802 standard; Peptide; 8 AA.
XX
AC   AAU76802;
XX
DT   21-MAY-2002  (first entry)
XX
DE   MHC class I-restricted epitope of chicken ovalbumin.
XX
KW   Chicken; ovalbumin; immunostimulant; T cell epitope; inosine; cytosine;
KW   polycationic peptide; systemic immune response; MHC class I; vaccine;
KW   major histocompatibility complex class I.
XX
OS   Gallus gallus.
XX
PN   WO200193903-A1.
XX
PD   13-DEC-2001.
XX
PF   07-JUN-2001; 2001WO-EP06437.
XX
PR   08-JUN-2000; 2000AT-0001000.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Egyed A,  Lingnau K,  Mattner F,  Buschle M,  Schmidt W;
XX
DR   WPI; 2002-205813/26.
XX
PT   Pharmaceutical composition for the preparation of vaccine comprises T
PT   cell epitope(s) or its mixture, polycationic peptide and nucleic acid
PT   based on inosine and cytosine  -
XX
PS   Example 1; Page 10; 45pp; English.
XX
CC   The invention relates to a pharmaceutical composition comprising a T cell
CC   epitope(s) or its mixture, a polycationic peptide and a nucleic acid
CC   based on inosine and cytosine. The composition of the invention induces a
CC   systemic immune response and is used for the preparation of a vaccine.
CC   This sequence represents an MHC class I-restricted epitope of chicken
CC   ovalbumin, used to test enhancement of immune response against an
CC   ovalbumin-derived peptide.
XX
SQ   Sequence   8 AA;

  Query Match            100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 59
AAU76869
ID   AAU76869 standard; Peptide; 8 AA.
XX
AC   AAU76869;
XX
DT   21-MAY-2002  (first entry)
XX
DE   OVA peptide fragment.
XX
KW   Membrane vesicle; exosome; density cushion centrifugation;
KW   dendritic cell; MHC; major histocompatibility complex; CD1; tumour;
KW   immunotherapy treatment; cancer; infection; immune disease; antitumour;
KW   cytostatic; OVA.
XX
OS   Unidentified.
XX
PN   WO200182958-A2.
XX
PD   08-NOV-2001.
XX
PF   11-APR-2001; 2001WO-EP04173.
XX
PR   27-APR-2000; 2000US-0561205.
PR   09-FEB-2001; 2001US-0780748.
XX
PA   (APCE-) AP CELLS INC.
XX
PI   Lamparski H,  Ruegg C,  Le Pecq J,  Hsu D,  Yao J;
XX
DR   WPI; 2002-066489/09.
XX
PT   Preparing membrane vesicle from biological sample for treating cancer,
PT   by culturing membrane vesicle-producing cells to release vesicles,
PT   enriching vesicles and subjecting sample to density cushion
PT   centrifugation  -
XX
PS   Example 12; Page 55; 103pp; English.
XX
CC   The invention relates to a method for preparing membrane vesicles (in
CC   particular exosomes) from a biological sample, comprising culturing a
CC   population of membrane vesicle-producing cells under conditions allowing
CC   the release of the vesicles, enriching the vesicles and treating the
CC   enriched biological sample by density cushion centrifugation. Immunogenic
CC   membrane vesicles are useful for producing an immune response in a
```

```
CC   subject, by obtaining a biological sample containing dendritic cells,
CC   isolating or purifying a membrane vesicle from the sample, contacting the
CC   purified vesicle with a peptide or a lipid under conditions allowing the
CC   peptide or lipid to bind an MHC or CD1 molecule at the surface of the
CC   vesicle, and administering the vesicle to the subject. Membrane vesicles
CC   are useful for immunotherapy treatment or prophylaxis of tumours, and for
CC   treating various disease conditions such as cancer, infections, and
CC   immune diseases. This sequence represents an OVA peptide fragment which
CC   is contacted with an exosome, to create an immunogenic membrane vesicle.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 60
AAU11866
ID   AAU11866 standard; Peptide; 8 AA.
XX
AC   AAU11866;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Ovalbumin drived control peptide.
XX
KW   T0 terminator; pGA; DNA vaccine; anti-HIV; virucide;
KW   Human Immunodeficiency Virus; HIV; Gag; HIV gp120; HIV Pol; HIV Env;
KW   HIV VLP; measles fusion protein; measles haemagglutinin; epitope;
KW   measles nucleoprotein; influenza haemagglutinin; C3d gene; ovalbumin;
KW   cell-mediated immune response; humoral immune response; infection.
XX
OS   Unidentified.
XX
PN   WO200192470-A2.
XX
PD   06-DEC-2001.
XX
PF   02-MAR-2001; 2001WO-US06795.
XX
PR   02-MAR-2000; 2000US-186364P.
PR   01-DEC-2000; 2000US-251083P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Robinson HL,  Smith JM,  Ross TM,  Bright RA,  Hua J,  Ellenberger D;
XX
DR   WPI; 2002-075465/10.
XX
PT   Novel pGA vector useful for immunising patient against measles,
PT   influenza has termination sequence encoding lambda T0 terminator and a
PT   eukaryotic transcription cassette with vaccine insert encoding
PT   immunogens of pathogens -
XX
PS   Example 14; Page 61; 174pp; English.
XX
CC   The invention relates to a vector (a pGA construct) comprising a
CC   termination sequence coding for the lambda T0 terminator, a prokaryotic
CC   origin of replication, a selectable marker gene and a eukaryotic
CC   transcription cassette comprising a vaccine insert encoding one or more
CC   immunogens derived from a pathogen e.g. Human Immunodeficiency Virus
CC   (HIV) Gag, HIV gp120, HIV Pol, HIV Env, HIV VLP, or its mutants, measles
CC   fusion protein, measles haemagglutinin, measles nucleoprotein, influenza
CC   haemagglutinin, or its mutants, or subsequences, and optionally at least
CC   one C3d gene, is useful for immunising or treating a patient, when
CC   administered by an intramuscular or intradermal route. The immunisation
CC   methods using pGA elicit both cell-mediated and humoral immune responses
CC   that may limit the infection, spread or growth of the pathogen and result
CC   in protection against subsequent challenge against the pathogen. The
CC   terminator sequence present prevents read-through from the kanamycin
CC   cassette into vaccine sequences while the plasmid is being produced in
CC   bacteria. Prevention of transcriptional read-through stabilises vaccine
CC   insert sequences by limiting the exposure of secondary structures that
CC   can be recognised by bacterial endonucleases. The present sequence
CC   is an ovalbumin control peptide used in an experiment to measure the
CC   T-cell response in monkeys inoculated with a pGA vector carrying vaccinia
CC   virus genes.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 61
AAU11239
ID   AAU11239 standard; peptide; 8 AA.
XX
AC   AAU11239;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Immunodominant Kb-restricted Cytotoxic T lymphocyte epitope #2.
XX
KW   Cytostatic; vaccine; tetanus toxin; FrC; tumour; CTL;
KW   cytotoxic T-lymphocyte; immunodominant Kb-restricted CTL epitope.
XX
OS   Unidentified.
XX
PN   WO200179510-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-GB01719.
XX
PR   17-APR-2000; 2000GB-0009470.
XX
PA   (CANC-) CANCER RES VENTURES LTD.
XX
PI   Rice J,  Stevenson F;
XX
DR   WPI; 2002-066370/09.
XX
PT   Nucleic acid construct, useful to immunise against various diseases
PT   including cancer, expresses the first domain of tetanus toxin FrC fused
PT   to a disease peptide antigen to provide a vaccine  -
XX
PS   Disclosure; Page 25; 71pp; English.
XX
CC   The invention relates to a nucleic acid construct for delivery into
CC   living cells in vivo, to induce an immune response to a disease peptide
CC   antigen, where the construct directs expression of a fusion protein
CC   comprising the peptide antigen and the first domain of FrC. Also
CC   included are a nucleic acid vector comprising the above construct,
CC   a host cell comprising the above construct or vector and a method of
CC   producing a nucleic acid construct for inducing an immune response.
CC   The method comprises identifying a nucleic acid sequence encoding a
CC   disease peptide antigen comprising epitopes characteristic of the
CC   disease, cloning the nucleic acid sequence, introducing the cloned
CC   nucleic acid into a vector which allows the antigen to be expressed as a
CC   fusion with a first domain FrC from tetanus toxin, and optionally
CC   isolating the construct from the vector. The construct or vector is used
CC   as a vaccine to induce an immune response, particularly to tumour
CC   antigens. The present sequence is an immunodominant Kb-restricted
CC   cytotoxic T-lymphocyte (CTL) epitope suitable for inclusion in the
CC   vaccine of the invention.
```

```
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 62
AAU09820
ID   AAU09820 standard; peptide; 8 AA.
XX
AC   AAU09820;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Ovalbumin-derived class I H-2Kb restricted peptide (245).
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide (245); vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Unidentified.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
PS   Example 1; Page 8; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents ovalbumin-derived class I H-2Kb restricted peptide
CC   (245) used to produce the modified hydrophobic peptides described in
CC   the method of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 63
AAE13215
ID   AAE13215 standard; peptide; 8 AA.
XX
AC   AAE13215;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Hen egg ovalbumin (OVA) peptide.
XX
KW   Cytotoxic T lymphocyte; CTL; T cell; tumour load; cancer radiotherapy;
KW   immunostimulatory sequence oligonucleotide; ISS-ODN; chemotherapy;
KW   immunosuppression; transplantation; autoimmune disease; infection;
KW   acquired immune deficiency syndrome; AIDS; intracellular pathogen;
KW   cytomegalovirus; mycobacterial infection; Epstein-Barr virus;
KW   varicella zoster virus; human immunodeficiency virus; HIV; hen;
KW   ovalbumin; OVA.
XX
OS   Unidentified.
XX
PN   WO200172123-A1.
XX
PD   04-OCT-2001.
XX
PF   28-MAR-2001; 2001WO-US10118.
XX
PR   28-MAR-2000; 2000US-192537P.
PR   11-MAY-2000; 2000US-203567P.
PR   05-JUL-2000; 2000US-215895P.
XX
PA   (REGC ) UNIV CALIFORNIA.
PA   (VETE-) DEPT VETERANS AFFAIRS.
XX
PI   Raz E,  Cho HJ,  Richman DD,  Horner AA;
XX
DR   WPI; 2002-010699/01.
XX
PT   Increasing antigen-specific cytotoxic T lymphocyte activity in a CD4+ T
PT   cell deficient individual, useful to treat immunodeficiency and block
PT   HIV infection, comprises administering immunostimulatory nucleic acid
PT   -
XX
PS   Example 1; Page 44; 91pp; English.
XX
CC   The present invention relates to a method for increasing antigen-specific
CC   cytotoxic T lymphocyte (CTL) activity in a CD4+ T cell-deficient
CC   individual, comprising administering an immunostimulatory sequence
CC   oligonucleotide (ISS-ODN). The immunostimulatory nucleic acids of the
CC   invention are used in CD4+ T cell-deficient individuals to decrease
CC   tumour load, to treat a primary or acquired immunodeficiency,
CC   particularly where the acquired immunodeficiency is temporary and due
CC   to cancer radiotherapy or chemotherapy or immunosuppression following
CC   bone marrow or organ transplantation, or autoimmune disease treatment,
CC   or is acquired immunodeficiency syndrome (AIDS). The nucleic acids may
CC   be used to treat a person at risk of becoming CD4+ T cell-deficient,
CC   particularly where someone at risk of cancer recurrence. They are also
CC   used to treat infection, particularly by an intracellular pathogen,
CC   especially one caused by cytomegalovirus, Mycobacterium tuberculosis,
CC   M. avium, Epstein-Barr virus, a fungus yeast, varicella zoster virus or
CC   human immunodeficiency virus (HIV). The present sequence is a hen egg
CC   ovalbumin (OVA) peptide, used in the exemplification of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
```

```
              ||||||||
Db          1 SIINFEKL 8


RESULT 64
AAE13436
ID   AAE13436 standard; peptide; 8 AA.
XX
AC   AAE13436;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken ovalbumin major histocompatability complex class I epitope.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatibility complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Example; Page 15; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatibility complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken ovalbumin MHC class I epitope.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 65
AAU75056
ID   AAU75056 standard; peptide; 8 AA.
XX
AC   AAU75056;
XX
DT   23-APR-2002  (first entry)
XX
DE   Ovalbumin antigenic peptide.
XX
KW   Ovalbumin; Th1; Th2; antigen; immunosuppressive; cytostatic;
KW   adjuvant; lipopolysaccharide; lipid; vaccine; immunogenicity;
KW   immunocompetence; autoimmune disease; infectious disease; OVA;
KW   graft-versus- host disease; tumour; transgenic T cell; chicken.
XX
OS   Gallus gallus.
XX
PN   WO200197838-A2.
XX
PD   27-DEC-2001.
XX
PF   18-JUN-2001; 2001WO-US19411.
XX
PR   16-JUN-2000; 2000US-212182P.
XX
PA   (BAYU ) BAYLOR RES INST.
XX
PI   Pulendran B,  Banchereau JF,  Cutler CW;
XX
DR   WPI; 2002-114543/15.
XX
PT   Use of adjuvants comprising isolated lipid groups such as Porphyromonas
PT   gingivalis lipopolysaccharides or its detoxified forms or derivatives
PT   for preparation of compositions to elicit T-helper cell responses
PT   mammals -
XX
PS   Disclosure; Page 20; 58pp; English.
XX
CC   This invention relates to the use of adjuvants comprising isolated lipid
CC   groups such as Porphyromonas gingivalis lipopolysaccharide, its
CC   detoxified forms or derivatives, for preparing compositions for
CC   eliciting Th2 responses, enhancing vaccine immunogenicity, modulating
CC   immunocompetence, treating autoimmune/infectious disease, stimulating
CC   interleukin-5 (IL-5)/IL-13 production or dampening interferon gamma
CC   production in mammal. The adjuvant or a pharmaceutical compound
CC   containing it iS useful for enhancing antibody harvest in a laboratory
CC   animal through an elicited Th2 immune response, or by modulating Th2
CC   immune responses. The adjuvant is also useful to study the Th2
CC   immune response in laboratory animal research, for the treatment or
CC   prophylactic vaccination of humans or animals against graft-versus-
CC   host disease, and for treating tumours. The present sequence
CC   represents the Ovalbumin (OVA) peptide used as an antigen to
CC   stimulate transgenic T cells in the method of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 66
ABU07743
ID   ABU07743 standard; Peptide; 8 AA.
XX
AC   ABU07743;
XX
DT   23-MAY-2003  (first entry)
XX
DE   Chicken ovalbumin epitope presented on murine Kb MHC class I.
XX
KW   Chicken; ovalbumin; OVA; cytolysin; vaccine delivery; Kb MHC class I;
KW   intracellular delivery vehicle; nonvirulent bacterium; drug delivery;
KW   gene therapy; biosynthesis; high level protein delivery;
KW   major histocompatability complex; cytosolic protein delivery.
XX
OS   Gallus gallus.
XX
PN   US2002142007-A1.
```

```
XX
PD   03-OCT-2002.
XX
PF   07-SEP-2001; 2001US-0949109.
XX
PR   13-AUG-1998;   98US-0133914.
PR   21-DEC-1999;   99US-0469197.
XX
PA   (PORT/) PORTNOY D A.
PA   (HIGG/) HIGGINS D E.
XX
PI   Portnoy DA,  Higgins DE;
XX
DR   WPI; 2003-328328/31.
XX
PT   New nonvirulent bacterium with genes coding for a non-secreted foreign
PT   cytolysin or a different foreign agent, useful as an intracellular
PT   delivery vehicle for delivering, e.g. vaccines, drugs or genes for
PT   therapy to eukaryotic cells  -
XX
PS   Example; Page 6; 14pp; English.
XX
CC   The invention relates to a nonvirulent bacterium, which comprises a first
CC   gene encoding a non-secreted foreign cytolysin operably linked to a
CC   heterologous promoter and a second gene encoding a different foreign
CC   agent. The nonvirulent bacterium is useful as an intracellular delivery
CC   vehicle, particularly of agents to eukaryotic cells. The nonvirulent
CC   bacterium is particularly useful for delivering foreign agents for
CC   diagnosis, therapy (e.g. prophylactics such as vaccine, delivery of
CC   therapeutic drug, or gene therapy) or biosynthesis. The nonvirulent
CC   bacterium is also useful for delivering nucleic acids that provide
CC   templates for transcription or translation, or provide modulators of
CC   transcription and/or translation. No protein purification is required
CC   compared to prior art delivery systems. In addition, high levels of
CC   protein can be delivered to the cytosol of virtually any cell and the
CC   levels can be controlled through the use of inducible promoters. L.
CC   monocytogenes LLO (listeriolysin or cytolysin) was transformed in
CC   E.coli cells and used as a system to deliver chicken ovalbumin to the
CC   cytosol of macrophages. The present sequence represents the chicken
CC   ovalbumin epitope presented on murine Kb major histocompatability
CC   complex, MHC, class I.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 67
ABU08619
ID   ABU08619 standard; Peptide; 8 AA.
XX
AC   ABU08619;
XX
DT   23-MAY-2003  (first entry)
XX
DE   Ovalbumin (OVA) residues 257-264.
XX
KW   Cancer; tumour; antigen-presenting cell; APC; tumour cell conjugate;
KW   cytokine; interleukin; interferon; IFN alpha; IFNbeta; IFNgamma;
KW   tumour necrosis factor; TNF; transforming growth factor; TGF;
KW   granulocyte-macrophage colony stimulating factor; GM-CSF; vaccine;
KW   melanoma; kidney cancer; pulmonary carcinoma; hepatic carcinoma;
KW   mammary cancer; prostatic carcinoma; gastric carcinoma; leukaemia;
KW   ovalbumin-specific tetramer; (OVA)-specific tetramer; ovalbumin;
KW   OVA.
XX
OS   Gallus gallus.
XX
PN   US2002182194-A1.
XX
PD   05-DEC-2002.
XX
PF   25-MAR-2002; 2002US-0106173.
XX
PR   04-APR-2001; 2001CN-0105852.
XX
PA   (SHAN-) SHANGHAI BRILLIANCE BIOTECH INST.
XX
PI   Ju D,  Tao Q,  Ye D;
XX
DR   WPI; 2003-328591/31.
XX
PT   New antigen-presenting cell and tumor cell conjugates, where the
PT   antigen-presenting cell is modified by a cytokine gene, useful for the
PT   preparation of a medicine for the therapy of cancer or a vaccine for
PT   the prophylaxis of cancer -
XX
PS   Example 1; Page 4; 23pp; English.
XX
CC   The invention describes an antigen-presenting cell (APC)/tumour cell
CC   conjugate, where the APC is modified by a cytokine gene (interleukin
CC   (IL-2, IL-3, IL-4, IL-6, IL-12, IL-18, interferon (IFN) alpha, IFNbeta,
CC   IFNgamma, tumour necrosis factor (TNF), transforming growth factor (TGF)
CC   and/or granulocyte-macrophage colony stimulating factor (GM-CSF)). The
CC   antigen-presenting cell/tumour cell conjugate is useful for the
CC   preparation of a medicine for the therapy of cancer or a vaccine for the
CC   prophylaxis of cancer. The cancer includes melanoma, kidney cancer,
CC   pulmonary carcinoma, hepatic carcinoma, mammary cancer, prostatic
CC   carcinoma, gastric carcinoma and leukaemia. This is the amino acid
CC   sequence of Ovalbumin (OVA) residues 257-264 used in the creation of
CC   Ovalbumin (OVA)-specific tetramers used in the vaccine of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 68
ABP57401
ID   ABP57401 standard; peptide; 8 AA.
XX
AC   ABP57401;
XX
DT   23-APR-2003  (first entry)
XX
DE   Synthetic 8mer peptide.
XX
KW   Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
KW   Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW   virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
OS   Synthetic.
XX
PN   WO2003000899-A1.
XX
PD   03-JAN-2003.
XX
PF   20-JUN-2002; 2002WO-GB02829.
XX
PR   22-JUN-2001; 2001GB-0015382.
XX
PA   (UYBR-) UNIV BRISTOL.
XX
PI   Hirst TR;
```

```
XX
DR   WPI; 2003-175291/17.
XX
PT   Use of a mutant form of B subunit of Escherichia coli heat labile
PT   enterotoxin or B subunit of cholera toxin for delivering an agent to a
PT   target cell for treating viral infection or cancer  -
XX
PS   Example 5; Page 45; 84pp; English.
XX
CC   The present invention describes a mutant form of B subunit of Escherichia
CC   coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
CC   from Vibrio cholerae which is useful for delivering an agent to a target
CC   cell, and has GM-1 ganglioside receptor binding activity but has reduced
CC   immunogenic and immunomodulatory activity relative to the wild-type form
CC   of EtxB or CtxB. Also described: (1) treating a disease or condition in
CC   a subject; (2) delivering the agent using the mutant to a target cell;
CC   (3) a composition; and (4) a kit for delivering the agent to a target
CC   cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
CC   can be used in vaccines. The mutant can be used for the preparation of
CC   a medicament for delivering an exogenous peptide, which is the agent,
CC   into the major histocompatibility complex (MHC) Class I antigen
CC   processing and presenting pathways to elicit a cytotoxic T lymphocyte
CC   (CTL) response, or for separate, simultaneous or combined use for
CC   treating viral infection or cancer. The mutant form of EtxB or CtxB
CC   enters mammalian cells without inducing a potent anti-B-subunit response
CC   and immunomodulatory response. It may be linked with an agent to
CC   upregulate the presentation of the antigen or antigenic determinant.
CC   The present sequence represents a peptide which is used in an example
CC   from the present invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 69
ABP58359
ID   ABP58359 standard; Peptide; 8 AA.
XX
AC   ABP58359;
XX
DT   07-APR-2003  (first entry)
XX
DE   Ovalbumin-derived peptide OVA257-264.
XX
KW   Chicken; ovalbumin; allergen; immunostimulant;
KW   oligodeoxynucleic acid; ODN; vaccine.
XX
OS   Gallus sp.
XX
PN   WO200295027-A2.
XX
PD   28-NOV-2002.
XX
PF   17-MAY-2002; 2002WO-EP05448.
XX
PR   21-MAY-2001; 2001AT-0000805.
XX
PA   (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Lingnau K,  Schellack C,  Schmidt W;
XX
DR   WPI; 2003-183880/18.
XX
PT   New oligodeoxynucleic acid molecules useful for the preparation of
PT   vaccine -
```

```
XX
PS   Example 3; Page 23; 57pp; English.
XX
CC   The present sequence is that of OVA257-264, a major
CC   histocompatibility complex class I (H-2Kb)-restricted epitope of
CC   chicken ovalbumin.  An example from the invention describes the
CC   generation of specific immune responses against this allergen-derived
CC   peptide using deoxyuridine monophosphate-modified oligonucleotide
CC   U-ODN 13 (see ABZ24776).  U-ODN 13 is an example of new
CC   oligodeoxynucleic acid (ODN) molecules useful in the preparation of
CC   vaccines.  The invention is based on the discovery that ODNs
CC   containing deoxyuridine residues have an immunostimulatory effect
CC   comparable to, or greater than, ODNs containing CpG motifs.
CC   Combining the ODN with an antigen strongly increases the potential
CC   of the antigen to raise the protection/immune response of a
CC   vaccinated individual.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 70
ABG73081
ID   ABG73081 standard; Peptide; 8 AA.
XX
AC   ABG73081;
XX
DT   02-APR-2003  (first entry)
XX
DE   MHC Class I peptide OVA-8.
XX
KW   Antigen-specific T lymphocyte; MHC-antigen complex; MHC Class I peptide;
KW   major histocompatibility complex; tumour-specific killer T cell;
KW   virus-specific killer T cell; cytostatic; virucide; OVA-8.
XX
OS   Synthetic.
XX
PN   US2002151690-A1.
XX
PD   17-OCT-2002.
XX
PF   05-NOV-1999;   99US-0434965.
XX
PR   12-AUG-1997;   97US-0909549.
XX
PA   (LUXE/) LUXEMBURG A T.
PA   (JACK/) JACKSON M R.
PA   (PETE/) PETER P A.
XX
PI   Luxemburg AT,  Jackson MR,  Peter PA;
XX
DR   WPI; 2003-182532/18.
XX
PT   Enriching antigen-specific T lymphocytes, for purifying or expanding in
PT   vitro tumour- or virus-specific killer T cells for cell therapy,
PT   comprises capture of the lymphocytes on a substrate coated with
PT   antigenic peptide-MHC complexes  -
XX
PS   Example 2; Page 5; 40pp; English.
XX
CC   The invention relates to a method for enriching antigen-specific T
CC   lymphocytes, comprising contacting a heterogenous population of
CC   antigen-specific T lymphocytes with a matrix comprising MHC-antigen
CC   complexes for a period of time sufficient to allow the antigen-specific T
CC   lymphocytes to interact with the matrix, and eluting the antigen-specific
CC   T lymphocytes from the matrix to provide an enriched population of
```

```
CC   antigen-specific T lymphocytes. The MHC-antigen complexes comprise one or
CC   more antigens. Also claimed is a matrix for capturing antigen-specific T
CC   lymphocytes, comprising a support having on its surface an immobilised
CC   Class I peptide and a predetermined amount of an antigen, or for
CC   capturing antigens, comprising a support having on its surface an
CC   immobilised empty Class I peptide which is capable of binding one or more
CC   antigens, and isolating antigen-specific T lymphocytes from a
CC   heterogenous population of cells from a patient. The methods are useful
CC   for enriching antigen-specific T lymphocytes to purify and expand in
CC   vitro tumour- and virus-specific killer T cells for cell therapy. The
CC   methods are also useful for isolating or preparing a population of
CC   antigen-specific T lymphocytes from a patient for treatment of the
CC   patient's disease or condition. This sequence represents an MHC Class I
CC   peptide used in the method of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 71
ABP56760
ID   ABP56760 standard; peptide; 8 AA.
XX
AC   ABP56760;
XX
DT   31-MAR-2003  (first entry)
XX
DE   Ovalbumin derived peptide OVA 257-264.
XX
KW   Stabilisation; polycationic polymer; medicine; vaccination;
KW   gene therapy; drug; ovalbumin.
XX
OS   Gallus sp.
OS   Synthetic.
XX
PN   WO200294845-A2.
XX
PD   28-NOV-2002.
XX
PF   17-MAY-2002; 2002WO-EP05447.
XX
PR   21-MAY-2001; 2001AT-0000805.
XX
PA   (INTE-) INTERCELL BIOMEDIZINISCHE FORSCHUNGS.
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Schellack C,  Lingnau K,  Schmidt W;
XX
DR   WPI; 2003-140356/13.
XX
PT   Use of polycationic polymer for stabilization of nucleic acids  -
XX
PS   Example 1; Page 6; 28pp; English.
XX
CC   The present invention describes a method for the stabilisation of nucleic
CC   acids involving contacting nucleic acids with a polycationic polymer in
CC   aqueous solution or suspension. The method can be used for the
CC   stabilisation of nucleic acid which can be used in medicines e.g.
CC   vaccinations, as a general immunostimicrolitresant, as an antisense drug
CC   or gene therapy drug. The present sequence represents an ovalbumin
CC   derived peptide, designated OVA 257-264, which is used in an example
CC   from the present invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 72
ABP60027
ID   ABP60027 standard; Peptide; 8 AA.
XX
AC   ABP60027;
XX
DT   07-MAR-2003  (first entry)
XX
DE   Ovalbumin antigenic peptide.
XX
KW   TOP; thimet oligopeptidase; EC3.4.25.15; cytostatic; tumour;
KW   immunostimulant; major histocompatibility complex class I; MHC;
KW   T-cell immunity; chicken.
XX
OS   Gallus gallus.
XX
PN   WO200279388-A2.
XX
PD   10-OCT-2002.
XX
PF   01-APR-2002; 2002WO-US10385.
XX
PR   30-MAR-2001; 2001US-280669P.
XX
PA   (UYMA-) UNIV MASSACHUSETTS.
XX
PI   Rock KL,  Goldberg AL;
XX
DR   WPI; 2003-103265/09.
XX
PT   New recombinant cell comprising an exogenously derived nucleic acid
PT   coding for a thimet oligopeptidase polypeptide, useful for modulating
PT   an antigenic response in a mammal for treating e.g., tumour  -
XX
PS   Example 1; Page 50; 73pp; English.
XX
CC   The invention relates to a new recombinant cell comprising an exogenously
CC   derived nucleic acid that codes for a thimet oligopeptidase (TOP)
CC   polypeptide. The TOP polypeptide is overexpressed in the cell compared to
CC   a wild-type cell from which the recombinant cell is derived. The activity
CC   of TOP may be described as cytostatic and immunostimulatory. Thimet
CC   oligopeptidase (TOP, EC3.4.25.15) plays a key role in modulating levels
CC   of major histocompatibility complex (MHC) class I-presented peptides. The
CC   recombinant host cell of the invention is useful for modulating an
CC   antigenic response in a mammal. Methods of the invention are useful for
CC   screening a test compound for its ability to serve as an immunomodulatory
CC   agent and identifying an antigen resistant to thimet oligopeptidase
CC   degradation. A method of the invention is useful for increasing CD8
CC   T-cell immunity, which uses vaccination with a TOP inhibitor for
CC   decreasing TOP expression or activity. The vaccination method uses
CC   treated tumour cells, antigen bearing/pulsed dendritic cells or injection
CC   of a viral vector. The recombinant host cell is useful for treating
CC   tumours. The current sequence represents an ovalbumin antigenic peptide
CC   that is used in an example from the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8
```

```
RESULT 73
ABU11029
ID   ABU11029 standard; Peptide; 8 AA.
XX
AC   ABU11029;
XX
DT   04-FEB-2003  (first entry)
XX
DE   Ovalbumin immunodominant epitope.
XX
KW   Epitope; antigen-specific immunoglobulin; Ig; early/late promoter;
KW   heavy chain constant region; light chain constant region;
KW   variable region; camelised Ig heavy chain variable region; MHC; CTL;
KW   major histocompatibility class; cytotoxic T-lyphocyte.
XX
OS   Unidentified.
XX
PN   US2002123057-A1.
XX
PD   05-SEP-2002.
XX
PF   14-NOV-2001; 2001US-0987456.
XX
PR   17-NOV-2000; 2000US-249268P.
PR   18-JAN-2001; 2001US-262067P.
PR   27-FEB-2001; 2001US-271424P.
PR   15-JUN-2001; 2001US-298087P.
XX
PA   (UYRP ) UNIV ROCHESTER.
XX
PI   Zauderer M,  Smith ES;
XX
DR   WPI; 2003-066785/06.
XX
PT   Selecting polynucleotides which encode antigen-specific immunoglobulin
PT   molecules, by introducing the library of polynucleotides into the host
PT   cells, and recovering the polynucleotides of the library for the
PT   antigen -
XX
PS   Example 5; Page 45; 108pp; English.
XX
CC   The invention relates to selecting polynucleotides which encode antigen
CC   -specific immunoglobulins (Ig) (or fragments) comprising introducing into
CC   a population of host cells, a 1st and 2nd library of polynucleotides
CC   encoding, several 1st and 2nd Ig subunit polypeptides, permitting
CC   expression of Ig molecules (via control element e.g. an early/late
CC   promoter), contacting Ig molecules with an antigen,
CC   and recovering polynucleotides of the 1st library for the antigen.
CC   The Ig molecules are heavy and light chain constant regions and
CC   variable regions linked via peptide linkers and optionally directed via
CC   signal peptides or transmembrane domains to different cell compartments.
CC   Also included is a method of selecting polynucleotides which encode a
CC   single-domain antigen-specific Ig molecule (its anti-specific fragment),
CC   by: (a) introducing into a population of eukaryotic host cells capable of
CC   expressing the Ig molecule a library of polynucleotides encoding
CC   (through operable association with a transcriptional control region)
CC   several single-domain Ig polypeptides (each comprising a Ig heavy chain
CC   constant region, a camelised Ig heavy chain variable region, and a
CC   signal peptide capable of directing cell surface expression or
CC   secretion of Ig subunit polypeptide); (b) permitting expression of Ig
CC   molecules (or antigen-specific fragments) from the host cells;
CC   (c) contacting the Ig molecules with an antigen; and (d) recovering
CC   polynucleotides of the library from those host cells expressing Ig
CC   molecules which bind the antigens. The methods are useful for selecting
CC   polynucleotides which encode an antigen-specific Ig molecule, or its
CC   fragment. The present sequence is a major histocompatibility class II
CC   (MHC II), cytotoxic T-lyphocyte (CTL) epitope expressed on the surface
CC   of host cells used in the method of the invention.
XX
SQ   Sequence   8 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 74
AAB84323
ID   AAB84323 standard; peptide; 9 AA.
XX
AC   AAB84323;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Peptide used to produce lemA peptide variants.
XX
XX
KW   lemA; CD8+ epitope; T cell response.
XX
OS   Synthetic.
XX
PN   WO200140275-A2.
XX
PD   07-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-US33027.
XX
PR   06-DEC-1999;   99US-0169227.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kurlander RJ,  Chao E,  Fields J;
XX
DR   WPI; 2001-389952/41.
XX
PT   New isolated variant of lemA, tlemA, comprising a hydrophobic element
PT   joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
PT   response or as a treatment or prophylactic against diseases  -
XX
PS   Disclosure; Page 7; 65pp; English.
XX
CC   The specification describes a peptide variant of lemA, comprising a
CC   hydrophobic element joined to a CD8+ epitope. The peptides may be
CC   used therapeutically by administering the peptides to a patient having
CC   a need to induce a directed CD8+ T cell response. The peptide may also
CC   be used as a preventive measure to avoid a disease or condition, or to
CC   treat subjects already afflicted with a disease. The present sequence
CC   was used to create peptides of the invention.
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          2 SIINFEKL 9


RESULT 75
ABP57402
ID   ABP57402 standard; peptide; 9 AA.
XX
AC   ABP57402;
XX
DT   23-APR-2003  (first entry)
XX
DE   Synthetic 9mer peptide.
XX
KW   Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
```

```
KW   Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW   virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
OS   Synthetic.
XX
PN   WO2003000899-A1.
XX
PD   03-JAN-2003.
XX
PF   20-JUN-2002; 2002WO-GB02829.
XX
PR   22-JUN-2001; 2001GB-0015382.
XX
PA   (UYBR-) UNIV BRISTOL.
XX
PI   Hirst TR;
XX
DR   WPI; 2003-175291/17.
XX
PT   Use of a mutant form of B subunit of Escherichia coli heat labile
PT   enterotoxin or B subunit of cholera toxin for delivering an agent to a
PT   target cell for treating viral infection or cancer  -
XX
PS   Example 5; Page 45; 84pp; English.
XX
CC   The present invention describes a mutant form of B subunit of Escherichia
CC   coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
CC   from Vibrio cholerae which is useful for delivering an agent to a target
CC   cell, and has GM-1 ganglioside receptor binding activity but has reduced
CC   immunogenic and immunomodulatory activity relative to the wild-type form
CC   of EtxB or CtxB. Also described: (1) treating a disease or condition in
CC   a subject; (2) delivering the agent using the mutant to a target cell;
CC   (3) a composition; and (4) a kit for delivering the agent to a target
CC   cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
CC   can be used in vaccines. The mutant can be used for the preparation of
CC   a medicament for delivering an exogenous peptide, which is the agent,
CC   into the major histocompatibility complex (MHC) Class I antigen
CC   processing and presenting pathways to elicit a cytotoxic T lymphocyte
CC   (CTL) response, or for separate, simultaneous or combined use for
CC   treating viral infection or cancer. The mutant form of EtxB or CtxB
CC   enters mammalian cells without inducing a potent anti-B-subunit response
CC   and immunomodulatory response. It may be linked with an agent to
CC   upregulate the presentation of the antigen or antigenic determinant.
CC   The present sequence represents a peptide which is used in an example
CC   from the present invention.
XX
SQ   Sequence   9 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 9.3e+05;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          2 SIINFEKL 9


RESULT 76
AAW04643
ID   AAW04643 standard; peptide; 10 AA.
XX
AC   AAW04643;
XX
DT   01-AUG-1997  (first entry)
XX
DE   Ovalbumin-derived activated CD8+ T cells epitope OVA10N.
XX
KW   Macrophage; artificial antigen presenting cell; APC; cancer;
KW   tumours; neoplasia; viral infection; retroviral infection;
KW   autoimmune.
XX
OS   Synthetic.
XX
PN   WO9637107-A1.
XX
PD   28-NOV-1996.
XX
PF   22-MAY-1996;   96WO-US07436.
XX
PR   23-MAY-1995;   95US-0447761.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   DeBruijn MLH,  Jackson MR,  Peterson PA;
XX
DR   WPI; 1997-020850/02.
XX
PT   Prodn. of activated CD8+ T cells directed to specific antigen  - can
PT   specifically kill target cells useful to treat, e.g. cancer
XX
PS   Example 1; Page 26; 84pp; English.
XX
CC   The method for the production of activated CD8+ T cells specifically
CC   directed towards a particular antigen involves affixing peptides
CC   corresponding to the particular antigen to an artificial support;
CC   contacting macrophages with the affixed peptides for a time sufficient
CC   for the peptides to be engulfed, and at least a portion of the peptides
CC   to be presented on the surface of the macrophage; and contacting
CC   unprimed CD8+ T cells with the peptide presenting macrophages for a
CC   time sufficient to activate the unprimed CD8+ T cells. The present
CC   sequence represents a peptide designated OVA10N which corresponds to
CC   ovalbumin, a Kb-restricted peptide antigen. This represents the optimal
CC   peptide with the addition of two amino acids at th eamino-terminus.
CC   Small extensions to the optimal peptide affect the affinity of the
CC   peptide for soluble class I molecules in vitro e.g. the addition
CC   of two amino acids to the amino-terminus lowers the affinity to Kb by
CC   76-fold compared to the optimal peptide; addition of two amino acids to
CC   the carboxy-terminus lowers the affinity by 4-fold. The method,
CC   macrophages and artificial antigen presenting cell, having a peptide
CC   corresponding to the particular antigen present on its surface and at
CC   least a portion of an artificial support in its interior, can be used to
CC   treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
CC   infection or autoimmune or autoimmune-type conditions) in patients via
CC   the specific killing of target cells.
XX
SQ   Sequence   10 AA;

  Query Match             100.0%;  Score 38;  DB 18;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.15;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          3 SIINFEKL 10


RESULT 77
AAW04644
ID   AAW04644 standard; peptide; 10 AA.
XX
AC   AAW04644;
XX
DT   01-AUG-1997  (first entry)
XX
DE   Ovalbumin-derived activated CD8+ T cells epitope OVA10C.
XX
KW   Macrophage; artificial antigen presenting cell; APC; cancer;
KW   tumours; neoplasia; viral infection; retroviral infection;
KW   autoimmune.
XX
OS   Synthetic.
XX
PN   WO9637107-A1.
XX
PD   28-NOV-1996.
XX
```

```
PF   22-MAY-1996;   96WO-US07436.
XX
PR   23-MAY-1995;   95US-0447761.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   DeBruijn MLH,  Jackson MR,  Peterson PA;
XX
DR   WPI; 1997-020850/02.
XX
PT   Prodn. of activated CD8+ T cells directed to specific antigen  - can
PT   specifically kill target cells useful to treat, e.g. cancer
XX
PS   Example 1; Page 26; 84pp; English.
XX
CC   The method for the production of activated CD8+ T cells specifically
CC   directed towards a particular antigen involves affixing peptides
CC   corresponding to the particular antigen to an artificial support;
CC   contacting macrophages with the affixed peptides for a time sufficient
CC   for the peptides to be engulfed, and at least a portion of the peptides
CC   to be presented on the surface of the macrophage; and contacting
CC   unprimed CD8+ T cells with the peptide presenting macrophages for a
CC   time sufficient to activate the unprimed CD8+ T cells. The present
CC   sequence represents a peptide designated OVA10C which corresponds to
CC   ovalbumin, a Kb-restricted peptide antigen. This represents the optimal
CC   peptide with the addition of two amino acids at the carboxy-terminus.
CC   Small extensions to the optimal peptide affect the affinity of the
CC   peptide for soluble class I molecules in vitro e.g. the addition
CC   of two amino acids to the amino-terminus lowers the affinity to Kb by
CC   76-fold compared to the optimal peptide; addition of two amino acids to
CC   the carboxy-terminus lowers the affinity by 4-fold. The method,
CC   macrophages and artificial antigen presenting cell, having a peptide
CC   corresponding to the particular antigen present on its surface and at
CC   least a portion of an artificial support in its interior, can be used to
CC   treat conditions (e.g. cancer, tumours, neoplasia, viral or retroviral
CC   infection or autoimmune or autoimmune-type conditions) in patients via
CC   the specific killing of target cells.
XX
SQ   Sequence   10 AA;

  Query Match             100.0%;  Score 38;  DB 18;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.15;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 78
AAU09821
ID   AAU09821 standard; peptide; 10 AA.
XX
AC   AAU09821;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Modified ovalbumin-derived class I H-2Kb restricted peptide #1.
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Synthetic.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
PS   Example 1; Page 9; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC   peptide #1 as described in the method of the invention.
XX
SQ   Sequence   10 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.15;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          3 SIINFEKL 10


RESULT 79
AAU09825
ID   AAU09825 standard; peptide; 10 AA.
XX
AC   AAU09825;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Modified ovalbumin-derived class I H-2Kb restricted peptide #5.
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Synthetic.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
```

```
PS   Example 2; Page 9; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC   peptide #5 as described in the method of the invention.
XX
SQ   Sequence   10 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 10;
  Best Local Similarity   100.0%;  Pred. No. 0.15;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          3 SIINFEKL 10


RESULT 80
AAW14122
ID   AAW14122 standard; peptide; 12 AA.
XX
AC   AAW14122;
XX
DT   20-OCT-1997  (first entry)
XX
DE   OVA protein derived MHC class I binding peptide.
XX
KW   Major histocompatibility complex; MHC; target; binding; tumour;
KW   cancer; neoplasia; LSTRA; EL-4; identification; detection; screening;
KW   tissue typing; Bcr-abl; IFV; influenza.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 12
FT                   /note= "biotinylated"
XX
PN   WO9641188-A1.
XX
PD   19-DEC-1996.
XX
PF   07-JUN-1996;   96WO-US09680.
XX
PR   07-JUN-1995;   95US-0485610.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Cheever MA,  Chen W;
XX
DR   WPI; 1997-108657/10.
XX
PT   Identifying major histocompatibility complex class I binding mols. -
PT   using peptide(s) having a core of 7-14 amino acids with extra amino
PT   acids and a reporter gp. at the N- or C-terminus, useful for tissue
PT   typing
XX
PS   Example 3; Page 23; 41pp; English.
XX
CC   AAW14122 is a biotinylated peptides derived from the OVA (ovalbumin)
CC   protein which can be obtained from either an LSTRA or EL-4 tumour
CC   of Balb/c mice. The peptides bind to MHC class I molecules. This
CC   is useful for tissue typing or for screening for molecules that
CC   interact with MHC class I molecules. MHC class I molecules can be
CC   identified using the peptides and also the peptides are useful in
CC   vaccines against disease and infection e.g. caused by viruses,
CC   bacteria or tumours.
XX
SQ   Sequence   12 AA;

  Query Match             100.0%;  Score 38;  DB 18;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 81
AAU09822
ID   AAU09822 standard; peptide; 12 AA.
XX
AC   AAU09822;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Modified ovalbumin-derived class I H-2Kb restricted peptide #2.
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Synthetic.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
PS   Example 1; Page 9; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC   peptide #2 as described in the method of the invention.
XX
SQ   Sequence   12 AA;
```

```
  Query Match             100.0%;  Score 38;  DB 23;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12


RESULT 82
AAU09826
ID   AAU09826 standard; peptide; 12 AA.
XX
AC   AAU09826;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Modified ovalbumin-derived class I H-2Kb restricted peptide #6.
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Synthetic.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
PS   Example 2; Page 9; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC   peptide #6 as described in the method of the invention.
XX
SQ   Sequence   12 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12


RESULT 83
AAU09827
ID   AAU09827 standard; peptide; 12 AA.
XX
AC   AAU09827;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Modified ovalbumin-derived class I H-2Kb restricted peptide #7.
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Synthetic.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
PS   Example 3; Page 10; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC   peptide #7 as described in the method of the invention.
XX
SQ   Sequence   12 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 12;
  Best Local Similarity   100.0%;  Pred. No. 0.18;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 84
ABB76049
ID   ABB76049 standard; Peptide; 14 AA.
XX
AC   ABB76049;
XX
DT   12-JUL-2002  (first entry)
XX
```

```
DE   Peptide insert in CACTE5-Cys-Ova.
XX
KW   Bordetella pertussis; adenylate cyclase; CyaA; adenylcyclase;
KW   vector; drug delivery; antigen delivery; cell targetting; CD11b;
KW   CACTES-Cys-Ova.
XX
OS   Synthetic.
XX
PN   EP1188446-A1.
XX
PD   20-MAR-2002.
XX
PF   15-SEP-2000; 2000EP-0402562.
XX
PR   15-SEP-2000; 2000EP-0402562.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CENT NAT RECH SCI.
XX
PI   Leclerc C,  Guermonprez P,  Ladant D,  Guiso N,  Khelef N;
XX
DR   WPI; 2002-354020/39.
XX
PT   Use of Bordetella adenylcyclase to make proteinaceous vector, useful
PT   for drug or antigen delivery, selectively targets cells that express
PT   CD11b  -
XX
PS   Example A; Page 10; 34pp; English.
XX
CC   The present sequence is a peptide that was introduced into the
CC   catalytic domain of a detoxified form of the adenylate cyclase
CC   (CyaA, or adenylcyclase) of Bordetella pertussis.  A recombinant
CC   detoxified CyaA toxin, CACTE5-Cys-Ova, harbouring a unique cysteine
CC   that was derived from the peptide insert, was produced.  The
CC   protein was labeled on its unique cysteine, and used to detect
CC   CyaA binding to neutrophils.  Experiments showed that CyaA binding
CC   to the surface of 3 myeloid cell lines of mouse or human origin, as
CC   well as to human neutrophils, was mainly mediated through the
CC   CD11b/CD18 integrin.  The invention relates to the novel use of
CC   Bordetella CyaA as a proteinaceous vector for targetting a
CC   molecule of interest to the surface CD11b-expressing cells,
CC   especially dendritic cells and neutrophils.  The molecule of
CC   interest is translocated in the cytosol to prime a cytotoxic T
CC   lymphocyte response.  In a preferred embodiment, a peptide is
CC   inserted into the catalytic domain of CyaA at a permissive site.
CC   The peptide may be an intracellular bacterial cell, tumour, viral,
CC   fungal or parasite cell antigen (all claimed).  Alternatively, a
CC   drug, especially an antiinflammatory, is chemically coupled to
CC   CyaA for drug delivery.
XX
SQ   Sequence   14 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12

RESULT 85
AAU09823
ID   AAU09823 standard; peptide; 14 AA.
XX
AC   AAU09823;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Modified ovalbumin-derived class I H-2Kb restricted peptide #3.
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Synthetic.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
PS   Example 1; Page 9; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC   peptide #3 as described in the method of the invention.
XX
SQ   Sequence   14 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          7 SIINFEKL 14

RESULT 86
AAU09828
ID   AAU09828 standard; peptide; 14 AA.
XX
AC   AAU09828;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Modified ovalbumin-derived class I H-2Kb restricted peptide #8.
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Synthetic.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
```

```
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
PS   Example 4; Page 11; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC   peptide #8 as described in the method of the invention.
XX
SQ   Sequence   14 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          7 SIINFEKL 14


RESULT 87
AAU09824
ID   AAU09824 standard; peptide; 15 AA.
XX
AC   AAU09824;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Modified ovalbumin-derived class I H-2Kb restricted peptide #4.
XX
KW   Ovalbumin-derived class I H-2Kb restricted peptide; vaccine;
KW   immunostimulant; immunogenic; interferon-gamma-producing T cell; IFN;
KW   popliteal lymph node; spleen; immune response; systemic response.
XX
OS   Synthetic.
XX
PN   WO200178767-A2.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-EP04313.
XX
PR   14-APR-2000; 2000AT-0000657.
XX
PA   (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI   Mattner F,  Zauner W,  Schmidt W,  Buschle M;
XX
DR   WPI; 2002-025970/03.
XX
PT   Pharmaceutical preparation for use as a potent vaccine for inducing an
PT   improved immune response in a mammal, comprises a modified peptide  -
XX
PS   Example 1; Page 9; 18pp; English.
XX
CC   The invention relates to a pharmaceutical preparation comprising a
CC   modified peptide, which induces an improved immune response in a mammal
CC   compared to the wild type peptide. The neutral peptide (SIINFEKL) (one
CC   negatively charged (Glu), one positively charged (Lys) amino acid) was
CC   rendered negative by adding (at the N-terminus) Glu Glu or Glu Asp Glu
CC   Asp, respectively. Results showed that the addition of 4 negatively-
CC   charged amino acids (EDED) at the N-terminus of peptide SIINFEKL makes
CC   this peptide (in combination with poly-L-arginine) able to induce a high
CC   amount of specific interferon (IFN)-gamma-producing T cells in the
CC   draining (popliteal) lymph node (local response) and in the spleen
CC   (systemic response). Thus, the addition of hydrophobic amino acids as
CC   well as the addition of negatively charged amino acids transforms the
CC   peptide SIINFEKL to a good inducer of specific T cells. The modified
CC   peptides of the pharmaceutical composition induce a stronger immune
CC   response in a mammal compared to wild type antigens. The present
CC   sequence represents modified ovalbumin-derived class I H-2Kb restricted
CC   peptide #4 as described in the method of the invention.
XX
SQ   Sequence   15 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 0.23;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          8 SIINFEKL 15


RESULT 88
ABP57403
ID   ABP57403 standard; peptide; 16 AA.
XX
AC   ABP57403;
XX
DT   23-APR-2003  (first entry)
XX
DE   Synthetic 16mer peptide.
XX
KW   Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
KW   Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW   virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
OS   Synthetic.
XX
PN   WO2003000899-A1.
XX
PD   03-JAN-2003.
XX
PF   20-JUN-2002; 2002WO-GB02829.
XX
PR   22-JUN-2001; 2001GB-0015382.
XX
PA   (UYBR-) UNIV BRISTOL.
XX
PI   Hirst TR;
XX
DR   WPI; 2003-175291/17.
XX
PT   Use of a mutant form of B subunit of Escherichia coli heat labile
PT   enterotoxin or B subunit of cholera toxin for delivering an agent to a
PT   target cell for treating viral infection or cancer  -
XX
PS   Example 5; Page 45; 84pp; English.
XX
CC   The present invention describes a mutant form of B subunit of Escherichia
CC   coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
CC   from Vibrio cholerae which is useful for delivering an agent to a target
```

```
CC   cell, and has GM-1 ganglioside receptor binding activity but has reduced
CC   immunogenic and immunomodulatory activity relative to the wild-type form
CC   of EtxB or CtxB. Also described: (1) treating a disease or condition in
CC   a subject; (2) delivering the agent using the mutant to a target cell;
CC   (3) a composition; and (4) a kit for delivering the agent to a target
CC   cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
CC   can be used in vaccines. The mutant can be used for the preparation of
CC   a medicament for delivering an exogenous peptide, which is the agent,
CC   into the major histocompatibility complex (MHC) Class I antigen
CC   processing and presenting pathways to elicit a cytotoxic T lymphocyte
CC   (CTL) response, or for separate, simultaneous or combined use for
CC   treating viral infection or cancer. The mutant form of EtxB or CtxB
CC   enters mammalian cells without inducing a potent anti-B-subunit response
CC   and immunomodulatory response. It may be linked with an agent to
CC   upregulate the presentation of the antigen or antigenic determinant.
CC   The present sequence represents a peptide which is used in an example
CC   from the present invention.
XX
SQ   Sequence   16 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 16;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          9 SIINFEKL 16


RESULT 89
AAW19957
ID   AAW19957 standard; Peptide; 19 AA.
XX
AC   AAW19957;
XX
DT   10-NOV-1997  (first entry)
XX
DE   BiP-binding domain-OVA hybrid peptide.
XX
KW   Vaccine; immunotherapy; heat shock protein; BiP; OVA; cancer;
KW   infectious disease.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..8
FT                   /label= BiP
FT   Peptide         9..11
FT                   /label= Linker
FT   Peptide         12..19
FT                   /label= Ova
XX
PN   WO9706821-A1.
XX
PD   27-FEB-1997.
XX
PF   16-AUG-1996;   96WO-US13363.
XX
PR   18-AUG-1995;   95US-0002490.
PR   18-AUG-1995;   95US-0002479.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Hartl FU,  Hoe MH,  Houghton A,  Mayhew M,  Rothman JE;
PI   Takeuchi Y;
XX
DR   WPI; 1997-165035/15.
XX
PT   Compsn. for inducing immune response contg. antigen and heat shock
PT   protein - also new hybrid peptide and related nucleic acid, for
PT   treatment of infectious diseases and tumours
XX
PS   Example 1; Page 18; 58pp; English.
XX
CC   Hybrid peptides OVA-BiP (AAW19956) and BiP-OVA (AAW19957) comprise
CC   chicken OVA-peptide (see AAW19955) joined via a peptide linker to
CC   heat shock protein (HSP) BiP binding domain (see also AAW19951).
CC   The hybrid peptide can be combined in vitro with a HSP, such as
CC   hsp70, to form a complex that, when administered to a subject,
CC   induces an immune response.
XX
SQ   Sequence   19 AA;

  Query Match             100.0%;  Score 38;  DB 18;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 0.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         12 SIINFEKL 19


RESULT 90
AAW19956
ID   AAW19956 standard; Peptide; 19 AA.
XX
AC   AAW19956;
XX
DT   10-NOV-1997  (first entry)
XX
DE   OVA-BiP-binding domain hybrid peptide.
XX
KW   Vaccine; immunotherapy; heat shock protein; BiP; OVA; cancer;
KW   infectious disease.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..8
FT                   /label= OVA
FT   Peptide         9..11
FT                   /label= Linker
FT   Peptide         12..19
FT                   /label= BiP
XX
PN   WO9706821-A1.
XX
PD   27-FEB-1997.
XX
PF   16-AUG-1996;   96WO-US13363.
XX
PR   18-AUG-1995;   95US-0002490.
PR   18-AUG-1995;   95US-0002479.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Hartl FU,  Hoe MH,  Houghton A,  Mayhew M,  Rothman JE;
PI   Takeuchi Y;
XX
DR   WPI; 1997-165035/15.
XX
PT   Compsn. for inducing immune response contg. antigen and heat shock
PT   protein - also new hybrid peptide and related nucleic acid, for
PT   treatment of infectious diseases and tumours
XX
PS   Example 1; Page 18; 58pp; English.
XX
CC   Hybrid peptides OVA-BiP (AAW19956) and BiP-OVA (AAW19957) comprise
CC   chicken OVA-peptide (see AAW19955) joined via a peptide linker to
CC   heat shock protein (HSP) BiP binding domain (see also AAW19951).
CC   The hybrid peptide can be combined in vitro with a HSP, such as
CC   hsp70, to form a complex that, when administered to a subject,
CC   induces an immune response.
XX
SQ   Sequence   19 AA;
```

```
  Query Match             100.0%;  Score 38;  DB 18;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 0.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 91
AAE13446
ID   AAE13446 standard; peptide; 19 AA.
XX
AC   AAE13446;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken MHC class I peptide antigen #1.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatability complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "MHC class I epitope"
FT   Region          9..11
FT                   /note= "Linker"
FT   Region          12..19
FT                   /note= "Javelin sequence"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 13; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatability complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken MHC class I peptide antigen.
XX
SQ   Sequence   19 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 0.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 92
AAE13447
ID   AAE13447 standard; peptide; 19 AA.
XX
AC   AAE13447;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken MHC class I peptide antigen #2.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatability complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "Javelin sequence"
FT   Region          9..11
FT                   /note= "Linker"
FT   Region          12..19
FT                   /note= "MHC class I epitope"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 13; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatability complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken MHC class I peptide antigen.
XX
SQ   Sequence   19 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 0.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         12 SIINFEKL 19


RESULT 93
ABP57404
ID   ABP57404 standard; peptide; 19 AA.
```

```
XX
AC   ABP57404;
XX
DT   23-APR-2003  (first entry)
XX
DE   Synthetic 19mer peptide.
XX
KW   Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
KW   Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW   virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
OS   Synthetic.
XX
PN   WO2003000899-A1.
XX
PD   03-JAN-2003.
XX
PF   20-JUN-2002; 2002WO-GB02829.
XX
PR   22-JUN-2001; 2001GB-0015382.
XX
PA   (UYBR-) UNIV BRISTOL.
XX
PI   Hirst TR;
XX
DR   WPI; 2003-175291/17.
XX
PT   Use of a mutant form of B subunit of Escherichia coli heat labile
PT   enterotoxin or B subunit of cholera toxin for delivering an agent to a
PT   target cell for treating viral infection or cancer  -
XX
PS   Example 5; Page 45; 84pp; English.
XX
CC   The present invention describes a mutant form of B subunit of Escherichia
CC   coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
CC   from Vibrio cholerae which is useful for delivering an agent to a target
CC   cell, and has GM-1 ganglioside receptor binding activity but has reduced
CC   immunogenic and immunomodulatory activity relative to the wild-type form
CC   of EtxB or CtxB. Also described: (1) treating a disease or condition in
CC   a subject; (2) delivering the agent using the mutant to a target cell;
CC   (3) a composition; and (4) a kit for delivering the agent to a target
CC   cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
CC   can be used in vaccines. The mutant can be used for the preparation of
CC   a medicament for delivering an exogenous peptide, which is the agent,
CC   into the major histocompatibility complex (MHC) Class I antigen
CC   processing and presenting pathways to elicit a cytotoxic T lymphocyte
CC   (CTL) response, or for separate, simultaneous or combined use for
CC   treating viral infection or cancer. The mutant form of EtxB or CtxB
CC   enters mammalian cells without inducing a potent anti-B-subunit response
CC   and immunomodulatory response. It may be linked with an agent to
CC   upregulate the presentation of the antigen or antigenic determinant.
CC   The present sequence represents a peptide which is used in an example
CC   from the present invention.
XX
SQ   Sequence   19 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 0.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         12 SIINFEKL 19

RESULT 94
AAR32294
ID   AAR32294 standard; protein; 24 AA.
XX
AC   AAR32294;
XX
DT   25-MAR-2003  (updated)
DT   31-MAY-1993  (first entry)
XX
DE   Sequence of synthetic peptide ova 253-276 which corresp. to AAs 253-
DE   276 of ovalbumin.
XX
KW   Cytotoxin T lymphocyte response; epitope; antigen.
XX
OS   Synthetic.
XX
PN   WO9301831-A1.
XX
PD   04-FEB-1993.
XX
PF   24-JUL-1992;   92WO-US06193.
XX
PR   25-JUL-1991;   91US-0735069.
XX
PA   (IDEC-) IDEC PHARM CORP.
XX
PI   Rastetter WH,  Raychaudhuri S;
XX
DR   WPI; 1993-058526/07.
XX
PT   New compsn. comprising an antigen and a formulation - to induce a
PT   cytotoxic T-lymphocyte response, useful for treating malaria,
PT   HIV, influenza, hepatitis, herpes, cancer, etc.
XX
PS   Disclosure; Page 19; 56pp; English.
XX
CC   Carbone and Bevan demonstrated that cytotoxic T-lymphocyte (CTL)
CC   induced in C57BL/6 mice by EG7-ova transfectant, and by
CC   cytoplasmically ova-loaded splenocytes recognise EL4 cells coated
CC   with the peptide ova 258-276. To determine whether soluble ovalbumin
CC   in AF induces similar CTL responses, spleen cells were prepared from
CC   immunised mice and stimulated in vitro with EG7-ova. The effectors
CC   were tested against El4 cells coated with the peptide ova  253-276
CC   or with a control peptide derived from myelin basic protein (MBP 84-
CC   102). The results demonstrate that ova-AF primed CTL with a similar
CC   specificity to those primed by transfectants, or by cytoplasmically
CC   loaded ova.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   24 AA;

  Query Match             100.0%;  Score 38;  DB 14;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12

RESULT 95
AAR41450
ID   AAR41450 standard; Protein; 24 AA.
XX
AC   AAR41450;
XX
DT   25-MAR-2003  (updated)
DT   23-FEB-1994  (first entry)
XX
DE   Antigenic peptide bound by MHC class one molecules.
XX
KW   HLA; Human Leucocyte Antigen; MHC; Class one molecules; cancer;
KW   autoimmunity; transplant rejection; T-cell activation.
XX
OS   Synthetic.
XX
PN   WO9317095-A1.
XX
PD   02-SEP-1993.
XX
PF   18-FEB-1993;   93WO-US01557.
```

```
XX
PR   19-FEB-1992;   92US-0841662.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   Jackson M,  Langlade-demoyen P,  Peterson PA;
XX
DR   WPI; 1993-288401/36.
XX
PT   Prodn. and use of human class I MHC molecules for activation of
PT   CD8 cells - for therapy of e.g. cancer, viral, retroviral and
PT   auto-immune diseases
XX
PS   Disclosure; Page 77; 182pp; English.
XX
CC   Human class I MHC genes are inserted into a cell and placed under
CC   the control of an inducible promoter.  This provides a means of
CC   producing, loading and using Class I MHC molecules to specifically
CC   activate CD8 cells in vitro.  Activated cells can be used to
CC   specifically kill target cells and also to treat cancer as well as
CC   viral, retroviral, autoimmune and autoimmune-type diseases.  When
CC   conjugated to a toxin, empty human MHC molecules expressed by the
CC   cells can be used to inhibit transplant rejection.  A number of
CC   antigenic peptides (AAR41450-R41463) are synthesised to be bound by
CC   the MHC molecules and this binding can then activate the CD8 cells.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence   24 AA;

  Query Match             100.0%;  Score 38;  DB 14;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12

RESULT 96
AAW04645
ID   AAW04645 standard; peptide; 24 AA.
XX
AC   AAW04645;
XX
DT   01-AUG-1997  (first entry)
XX
DE   Ovalbumin-derived activated CD8+ T cells epitope OVA24.
XX
KW   Macrophage; artificial antigen presenting cell; APC; cancer;
KW   tumours; neoplasia; viral infection; retroviral infection;
KW   autoimmune.
XX
OS   Synthetic.
XX
PN   WO9637107-A1.
XX
PD   28-NOV-1996.
XX
PF   22-MAY-1996;   96WO-US07436.
XX
PR   23-MAY-1995;   95US-0447761.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   DeBruijn MLH,  Jackson MR,  Peterson PA;
XX
DR   WPI; 1997-020850/02.
XX
PT   Prodn. of activated CD8+ T cells directed to specific antigen  - can
PT   specifically kill target cells useful to treat, e.g. cancer
XX
PS   Example 1; Page 26; 84pp; English.
XX
CC   The method for the production of activated CD8+ T cells specifically
CC   directed towards a particular antigen involves affixing peptides
CC   corresponding to the particular antigen to an artificial support;
CC   contacting macrophages with the affixed peptides for a time sufficient
CC   for the peptides to be engulfed, and at least a portion of the peptides
CC   to be presented on the surface of the macrophage; and contacting
CC   unprimed CD8+ T cells with the peptide presenting macrophages for a
CC   time sufficient to activate the unprimed CD8+ T cells. The present
CC   sequence represents a peptide designated OVA24 which corresponds to
CC   ovalbumin, a Kb-restricted peptide antigen. This is not as efficient as
CC   the optimal peptide. The method, macrophages and artificial antigen
CC   presenting cell, having a peptide corresponding to the particular
CC   antigen present on its surface and at least a portion of an artificial
CC   support in its interior, can be used to treat conditions (e.g. cancer,
CC   tumours, neoplasia, viral or retroviral infection or autoimmune or
CC   autoimmune-type conditions) in patients via the specific killing of
CC   target cells.
XX
SQ   Sequence   24 AA;

  Query Match             100.0%;  Score 38;  DB 18;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12

RESULT 97
AAG65170
ID   AAG65170 standard; peptide; 24 AA.
XX
AC   AAG65170;
XX
DT   15-OCT-2002  (first entry)
XX
DE   Ovalbumin based peptide.
XX
KW   Cytotoxic T lymphocyte response; CTL; antigen; viral infection;
KW   bacterial infection; cancer; parasitic infection; immune response;
KW   non-toxic.
XX
OS   Unidentified.
XX
PN   US6270769-B1.
XX
PD   07-AUG-2001.
XX
PF   24-MAY-1995;   95US-0449728.
XX
PR   24-JUL-1992;   92US-0919787.
PR   25-JUL-1991;   91US-0735069.
XX
PA   (IDEC-) IDEC PHARM CORP.
XX
PI   Raychaudhuri S,  Rastetter WH;
XX
DR   WPI; 2001-564234/63.
XX
PT   Induction of cytotoxic T-lymphocyte responses -
XX
PS   Disclosure; Column 17-18; 24pp; English.
XX
CC   The present invention relates to a method of treating viral, parasitic
CC   and bacterial infections and cancer in humans, by administering an
CC   antigen which causes a cytotoxic T-lymphocyte response. Said antigen does
CC   not contain an immunostimulatory element. The method can also be used in
CC   domesticated animals. The present sequence is a peptide used as an
CC   antigen in the exemplification of the invention.
XX
SQ   Sequence   24 AA;
```

```
  Query Match             100.0%;  Score 38;  DB 22;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12


RESULT 98
AAB74439
ID   AAB74439 standard; peptide; 24 AA.
XX
AC   AAB74439;
XX
DT   29-MAY-2001  (first entry)
XX
DE   Myelin basic protein amino acids 84-102 peptide.
XX
KW   Cytotoxic T-lymphocyte; CTL response; immunostimulation; infection;
KW   cancer.
XX
OS   Unidentified.
XX
PN   US6197311-B1.
XX
PD   06-MAR-2001.
XX
PF   17-FEB-1998;   98US-0024220.
XX
PR   07-JUN-1995;   95US-0476674.
PR   25-JUL-1991;   91US-0735069.
PR   24-JUL-1992;   92US-0919787.
PR   07-DEC-1994;   94US-0351001.
XX
PA   (IDEC-) IDEC PHARM CORP.
XX
PI   Raychaudhuri S,  Rastetter WH,  Black A;
XX
DR   WPI; 2001-256350/26.
XX
PT   Treating papillomavirus-related tumor or malignancy, involves
PT   administering an antigen formulation substantially free of
PT   immunostimulatory peptides, and comprising human papillomavirus antigen
PT   and microfluidized adjuvant -
XX
PS   Disclosure; Column 10-11; 22pp; English.
XX
CC   The present invention describes a method of treating a
CC   papillomavirus-related tumour, involving administering an antigen
CC   formulation capable of inducing a cytotoxic T-lymphocyte (CTL) response
CC   specific to the papillomavirus antigen in the individual. This is useful
CC   in the treatment of cancer and infections, such as those due to HIV,
CC   bacteria, parasites, influenza, herpes virus and hepatitis viruses. The
CC   present sequence is a peptide used to demonstrate the method of the
CC   invention.
XX
SQ   Sequence   24 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12


RESULT 99
ABG31664
ID   ABG31664 standard; Peptide; 24 AA.
XX
AC   ABG31664;
XX
DT   05-NOV-2002  (first entry)
XX
DE   Ovalbumin (OVA) peptide fragment.
XX
KW   Ovalbumin; microfluidised antigen; detergent; micelle-forming agent;
KW   biodegradable oil; biocompatible oil; cytotoxic T lymphocyte; HIV;
KW   human immunodeficiency virus; herpes virus; malaria; influenza; cancer;
KW   hepatitis; respiratory syncytial virus; domesticated animal; OVA;
KW   agricultural animal.
XX
OS   Unidentified.
XX
PN   US2002039582-A1.
XX
PD   04-APR-2002.
XX
PF   20-DEC-2000; 2000US-0740003.
XX
PR   17-FEB-1998;   98US-0024220.
PR   25-JUL-1991;   91US-0735069.
PR   07-DEC-1994;   94US-0351001.
PR   07-JUN-1995;   95US-0476674.
PR   29-AUG-1997;   97US-0919787.
XX
PA   (IDEC-) IDEC PHARM CORP.
XX
PI   Raychaudhuri S,  Rastetter WH,  Black A;
XX
DR   WPI; 2002-607062/65.
XX
PT   Composition useful for inducing cytotoxic T lymphocyte response in
PT   domesticated animals and  humans comprises antigen mixed with
PT   microfluidised antigen formulation which is substantially free of
PT   immunostimulating peptides  -
XX
PS   Disclosure; Page 6; 31pp; English.
XX
CC   The invention relates to a composition comprising an antigen mixed with
CC   microfluidised antigen formulation comprising a stabilised detergent, a
CC   micelle-forming agent and a biodegradable/biocompatible oil. The
CC   composition is formulated as a stable oil-in-water emulsion substantially
CC   free of or lacking immunostimulating peptides, and is capable of inducing
CC   specific cytotoxic T lymphocyte response against antigens in vivo. The
CC   composition is useful for treating patients infected with human
CC   immunodeficiency virus (HIV) or herpes virus, and patients suffering from
CC   malaria, influenza, hepatitis, cancer or respiratory syncytial virus by
CC   administering a composition comprising HIV antigen, malaria-associated
CC   antigen, hepatitis-associated antigen, cancer-associated antigen, herpes
CC   antigen or respiratory syncytial antigen respectively, mixed with
CC   microfluidised antigen consisting essentially of two of stabilising
CC   detergent, micelle-forming agent and biodegradable and biocompatible oil,
CC   the antigen formulation being formulated as stable oil-in-water emulsion,
CC   and inducing cytotoxic T lymphocyte response in a patient e.g. human,
CC   domesticated animal or agricultural animal. This sequence represents an
CC   ovalbumin (OVA) peptide fragment used in the scope of the invention.
XX
SQ   Sequence   24 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 0.38;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          5 SIINFEKL 12


RESULT 100
ABP57405
ID   ABP57405 standard; peptide; 26 AA.
XX
AC   ABP57405;
```

```
XX
DT   23-APR-2003  (first entry)
XX
DE   Synthetic 26mer peptide.
XX
KW   Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
KW   Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW   virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
OS   Synthetic.
XX
PN   WO2003000899-A1.
XX
PD   03-JAN-2003.
XX
PF   20-JUN-2002; 2002WO-GB02829.
XX
PR   22-JUN-2001; 2001GB-0015382.
XX
PA   (UYBR-) UNIV BRISTOL.
XX
PI   Hirst TR;
XX
DR   WPI; 2003-175291/17.
XX
PT   Use of a mutant form of B subunit of Escherichia coli heat labile
PT   enterotoxin or B subunit of cholera toxin for delivering an agent to a
PT   target cell for treating viral infection or cancer  -
XX
PS   Example 5; Page 45; 84pp; English.
XX
CC   The present invention describes a mutant form of B subunit of Escherichia
CC   coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
CC   from Vibrio cholerae which is useful for delivering an agent to a target
CC   cell, and has GM-1 ganglioside receptor binding activity but has reduced
CC   immunogenic and immunomodulatory activity relative to the wild-type form
CC   of EtxB or CtxB. Also described: (1) treating a disease or condition in
CC   a subject; (2) delivering the agent using the mutant to a target cell;
CC   (3) a composition; and (4) a kit for delivering the agent to a target
CC   cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
CC   can be used in vaccines. The mutant can be used for the preparation of
CC   a medicament for delivering an exogenous peptide, which is the agent,
CC   into the major histocompatibility complex (MHC) Class I antigen
CC   processing and presenting pathways to elicit a cytotoxic T lymphocyte
CC   (CTL) response, or for separate, simultaneous or combined use for
CC   treating viral infection or cancer. The mutant form of EtxB or CtxB
CC   enters mammalian cells without inducing a potent anti-B-subunit response
CC   and immunomodulatory response. It may be linked with an agent to
CC   upregulate the presentation of the antigen or antigenic determinant.
CC   The present sequence represents a peptide which is used in an example
CC   from the present invention.
XX
SQ   Sequence   26 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 0.41;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         19 SIINFEKL 26


RESULT 101
ABP57406
ID   ABP57406 standard; peptide; 26 AA.
XX
AC   ABP57406;
XX
DT   23-APR-2003  (first entry)
XX
DE   Synthetic 26mer* peptide.
XX
KW   Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
KW   Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW   virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
OS   Synthetic.
XX
PN   WO2003000899-A1.
XX
PD   03-JAN-2003.
XX
PF   20-JUN-2002; 2002WO-GB02829.
XX
PR   22-JUN-2001; 2001GB-0015382.
XX
PA   (UYBR-) UNIV BRISTOL.
XX
PI   Hirst TR;
XX
DR   WPI; 2003-175291/17.
XX
PT   Use of a mutant form of B subunit of Escherichia coli heat labile
PT   enterotoxin or B subunit of cholera toxin for delivering an agent to a
PT   target cell for treating viral infection or cancer  -
XX
PS   Example 5; Page 45; 84pp; English.
XX
CC   The present invention describes a mutant form of B subunit of Escherichia
CC   coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
CC   from Vibrio cholerae which is useful for delivering an agent to a target
CC   cell, and has GM-1 ganglioside receptor binding activity but has reduced
CC   immunogenic and immunomodulatory activity relative to the wild-type form
CC   of EtxB or CtxB. Also described: (1) treating a disease or condition in
CC   a subject; (2) delivering the agent using the mutant to a target cell;
CC   (3) a composition; and (4) a kit for delivering the agent to a target
CC   cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
CC   can be used in vaccines. The mutant can be used for the preparation of
CC   a medicament for delivering an exogenous peptide, which is the agent,
CC   into the major histocompatibility complex (MHC) Class I antigen
CC   processing and presenting pathways to elicit a cytotoxic T lymphocyte
CC   (CTL) response, or for separate, simultaneous or combined use for
CC   treating viral infection or cancer. The mutant form of EtxB or CtxB
CC   enters mammalian cells without inducing a potent anti-B-subunit response
CC   and immunomodulatory response. It may be linked with an agent to
CC   upregulate the presentation of the antigen or antigenic determinant.
CC   The present sequence represents a peptide which is used in an example
CC   from the present invention.
XX
SQ   Sequence   26 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 0.41;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         19 SIINFEKL 26


RESULT 102
AAE13448
ID   AAE13448 standard; peptide; 30 AA.
XX
AC   AAE13448;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken MHC class I peptide antigen #3.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatability complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
```

```
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "Javelin sequence"
FT   Region          9..11
FT                   /note= "Linker"
FT   Region          12..19
FT                   /note= "MHC class I epitope"
FT   Region          20..22
FT                   /note= "Linker"
FT   Region          23..30
FT                   /note= "Javelin sequence"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 13; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatability complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken MHC class I peptide antigen.
XX
SQ   Sequence   30 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 30;
  Best Local Similarity   100.0%;  Pred. No. 0.48;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         12 SIINFEKL 19


RESULT 103
ABP57407
ID   ABP57407 standard; peptide; 31 AA.
XX
AC   ABP57407;
XX
DT   23-APR-2003  (first entry)
XX
DE   Synthetic 31mer peptide.
XX
KW   Escherichia coli heat labile enterotoxin; Etx; cholera toxin; Ctx;
KW   Vibrio cholerae; mutant; GM-1 ganglioside receptor; carrier; toxin;
KW   virucide; cytostatic; vaccine; viral infection; cancer; EtxB; CtxB.
XX
OS   Synthetic.
XX
PN   WO2003000899-A1.
XX
PD   03-JAN-2003.
XX
PF   20-JUN-2002; 2002WO-GB02829.
XX
PR   22-JUN-2001; 2001GB-0015382.
XX
PA   (UYBR-) UNIV BRISTOL.
XX
PI   Hirst TR;
XX
DR   WPI; 2003-175291/17.
XX
PT   Use of a mutant form of B subunit of Escherichia coli heat labile
PT   enterotoxin or B subunit of cholera toxin for delivering an agent to a
PT   target cell for treating viral infection or cancer  -
XX
PS   Example 5; Page 45; 84pp; English.
XX
CC   The present invention describes a mutant form of B subunit of Escherichia
CC   coli heat labile enterotoxin (EtxB) or B subunit of cholera toxin (CtxB)
CC   from Vibrio cholerae which is useful for delivering an agent to a target
CC   cell, and has GM-1 ganglioside receptor binding activity but has reduced
CC   immunogenic and immunomodulatory activity relative to the wild-type form
CC   of EtxB or CtxB. Also described: (1) treating a disease or condition in
CC   a subject; (2) delivering the agent using the mutant to a target cell;
CC   (3) a composition; and (4) a kit for delivering the agent to a target
CC   cell. Mutant EtxB and CtxB have virucide and cytostatic activities and
CC   can be used in vaccines. The mutant can be used for the preparation of
CC   a medicament for delivering an exogenous peptide, which is the agent,
CC   into the major histocompatibility complex (MHC) Class I antigen
CC   processing and presenting pathways to elicit a cytotoxic T lymphocyte
CC   (CTL) response, or for separate, simultaneous or combined use for
CC   treating viral infection or cancer. The mutant form of EtxB or CtxB
CC   enters mammalian cells without inducing a potent anti-B-subunit response
CC   and immunomodulatory response. It may be linked with an agent to
CC   upregulate the presentation of the antigen or antigenic determinant.
CC   The present sequence represents a peptide which is used in an example
CC   from the present invention.
XX
SQ   Sequence   31 AA;

  Query Match             100.0%;  Score 38;  DB 24;  Length 31;
  Best Local Similarity   100.0%;  Pred. No. 0.49;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         19 SIINFEKL 26


RESULT 104
AAW04646
ID   AAW04646 standard; peptide; 35 AA.
XX
AC   AAW04646;
XX
DT   01-AUG-1997  (first entry)
XX
DE   Ovalbumin-derived activated CD8+ T cells epitope OVA35.
XX
KW   Macrophage; artificial antigen presenting cell; APC; cancer;
KW   tumours; neoplasia; viral infection; retroviral infection;
KW   autoimmune.
XX
OS   Synthetic.
XX
PN   WO9637107-A1.
XX
PD   28-NOV-1996.
XX
PF   22-MAY-1996;   96WO-US07436.
XX
```

```
PR   23-MAY-1995;   95US-0447761.
XX
PA   (SCRI ) SCRIPPS RES INST.
XX
PI   DeBruijn MLH,  Jackson MR,  Peterson PA;
XX
DR   WPI; 1997-020850/02.
XX
PT   Prodn. of activated CD8+ T cells directed to specific antigen  - can
PT   specifically kill target cells useful to treat, e.g. cancer
XX
PS   Example 1; Page 26; 84pp; English.
XX
CC   The method for the production of activated CD8+ T cells specifically
CC   directed towards a particular antigen involves affixing peptides
CC   corresponding to the particular antigen to an artificial support;
CC   contacting macrophages with the affixed peptides for a time sufficient
CC   for the peptides to be engulfed, and at least a portion of the peptides
CC   to be presented on the surface of the macrophage; and contacting
CC   unprimed CD8+ T cells with the peptide presenting macrophages for a
CC   time sufficient to activate the unprimed CD8+ T cells. The present
CC   sequence represents a peptide designated OVA35 which corresponds to
CC   ovalbumin, a Kb-restricted peptide antigen. This is not as efficient as
CC   the optimal peptide. The method, macrophages and artificial antigen
CC   presenting cell, having a peptide corresponding to the particular
CC   antigen present on its surface and at least a portion of an artificial
CC   support in its interior, can be used to treat conditions (e.g. cancer,
CC   tumours, neoplasia, viral or retroviral infection or autoimmune or
CC   autoimmune-type conditions) in patients via the specific killing of
CC   target cells.
XX
SQ   Sequence   35 AA;

  Query Match            100.0%;  Score 38;  DB 18;  Length 35;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         16 SIINFEKL 23


RESULT 105
AAO26741
ID   AAO26741 standard; Peptide; 36 AA.
XX
AC   AAO26741;
XX
DT   27-MAR-2003  (first entry)
XX
DE   Chicken ovalbumin class I and II carrying antigenic peptide.
XX
KW   Cytostatic; universal polypeptidic carrier; Gb3 receptor; cytotoxic drug;
KW   tumour cell; immunogenic; gene therapy; chicken.
XX
OS   Gallus sp.
XX
PN   EP1229045-A1.
XX
PD   07-AUG-2002.
XX
PF   01-FEB-2001; 2001EP-0400255.
XX
PR   01-FEB-2001; 2001EP-0400255.
XX
PA   (CURI-) INST CURIE.
PA   (CNRS ) CENT NAT RECH SCI.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (UYPA-) UNIV CURIE PARIS VI P & M.
XX
PI   Johannes L,  Tartour E,  Goud B,  Fridman WH;
XX
DR   WPI; 2003-185783/19.
XX
PT   Novel universal polypeptidic carrier for targeting a molecule directly
PT   or indirectly to Gb3 receptor expressing cells, useful for directing
PT   cytotoxic drugs to tumor cells expressing Gb3 receptor  -
XX
PS   Example 3; Page 8; 24pp; English.
XX
CC   The invention relates to a universal polypeptidic carrier for targeting
CC   directly or indirectly a molecule to Gb3 receptor expressing cells. The
CC   universal polypeptidic carrier is useful for targeting directly or
CC   indirectly, to Gb3 receptor expressing cells, using molecules such as
CC   proteins, peptides, oligopeptides, glycoproteins, glycopeptides, nucleic
CC   acids, polynucleotides, or its combination, an antigen to be targeted to
CC   antigen presenting cells or a cytotoxic drug or pro-drug to be targeted
CC   to tumour cells expressing Gb3 receptor. The universal polypeptidic
CC   carrier is useful for delivering an expression vector containing a
CC   sequence of interest into a Gb3 receptor expressing cells which involves
CC   operably linking the expression vector to a lysine-rich peptide
CC   covalently linked to the Cys moiety of the universal polypeptidic
CC   carrier. The lysine rich peptide is preferably a 16-mer poly-lysine and
CC   the sequence of interest is preferably a sequence encoding an immunogenic
CC   peptide, or a sequence encoding a drug or a pro-drug becoming toxic for
CC   the Gb3 receptor expressing cells, or a sequence encoding a therapeutic
CC   active molecule. The universal polypeptidic carrier is useful for
CC   targeting a molecule to a Gb3 receptor expressing cell to enable the
CC   molecule to be internalised, processed and/or expressed in the cell
CC   expressing a Gb3 receptor. The universal polypeptidic carrier is useful
CC   as a carrier for introducing a nucleotide sequence in a target cell
CC   either for gene therapy or for obtaining recombinant cells expressing
CC   heterologous proteins. This sequence represents a chicken ovalbumin class
CC   I and II carrying antigenic peptide relating to the universal carrier
CC   protein of the invention.
XX
SQ   Sequence   36 AA;

  Query Match            100.0%;  Score 38;  DB 24;  Length 36;
  Best Local Similarity   100.0%;  Pred. No. 0.58;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          6 SIINFEKL 13


RESULT 106
AAB84325
ID   AAB84325 standard; peptide; 43 AA.
XX
AC   AAB84325;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Amino acid sequence of a lemS variant peptide.
XX
KW   lemA; CD8+ epitope; T cell response.
XX
OS   Synthetic.
XX
PN   WO200140275-A2.
XX
PD   07-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-US33027.
XX
PR   06-DEC-1999;   99US-0169227.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kurlander RJ,  Chao E,  Fields J;
XX
DR   WPI; 2001-389952/41.
XX
PT   New isolated variant of lemA, tlemA, comprising a hydrophobic element
```

```
PT   joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
PT   response or as a treatment or prophylactic against diseases  -
XX
PS   Disclosure; Page 17; 65pp; English.
XX
CC   The specification describes a peptide variant of lemA, comprising a
CC   hydrophobic element joined to a CD8+ epitope. The peptides may be
CC   used therapeutically by administering the peptides to a patient having
CC   a need to induce a directed CD8+ T cell response. The peptide may also
CC   be used as a preventive measure to avoid a disease or condition, or to
CC   treat subjects already afflicted with a disease. The present sequence
CC   represents a peptide of the invention.
XX
SQ   Sequence   43 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 43;
  Best Local Similarity   100.0%;  Pred. No. 0.69;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         28 SIINFEKL 35


RESULT 107
AAB84321
ID   AAB84321 standard; peptide; 47 AA.
XX
AC   AAB84321;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Amino acid sequence of a lemS variant peptide.
XX
KW   lemA; CD8+ epitope; T cell response.
XX
OS   Synthetic.
XX
PN   WO200140275-A2.
XX
PD   07-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-US33027.
XX
PR   06-DEC-1999;   99US-0169227.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kurlander RJ,  Chao E,  Fields J;
XX
DR   WPI; 2001-389952/41.
XX
PT   New isolated variant of lemA, tlemA, comprising a hydrophobic element
PT   joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
PT   response or as a treatment or prophylactic against diseases  -
XX
PS   Disclosure; Page 19; 65pp; English.
XX
CC   The specification describes a peptide variant of lemA, comprising a
CC   hydrophobic element joined to a CD8+ epitope. The peptides may be
CC   used therapeutically by administering the peptides to a patient having
CC   a need to induce a directed CD8+ T cell response. The peptide may also
CC   be used as a preventive measure to avoid a disease or condition, or to
CC   treat subjects already afflicted with a disease. The present sequence
CC   represents a peptide of the invention.
XX
SQ   Sequence   47 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 47;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         34 SIINFEKL 41


RESULT 108
AAB84322
ID   AAB84322 standard; peptide; 48 AA.
XX
AC   AAB84322;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Amino acid sequence of a SemA variant peptide.
XX
KW   lemA; CD8+ epitope; T cell response.
XX
OS   Synthetic.
XX
PN   WO200140275-A2.
XX
PD   07-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-US33027.
XX
PR   06-DEC-1999;   99US-0169227.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Kurlander RJ,  Chao E,  Fields J;
XX
DR   WPI; 2001-389952/41.
XX
PT   New isolated variant of lemA, tlemA, comprising a hydrophobic element
PT   joined to a CD8+ epitope, useful for inducing a directed CD8+ T cell
PT   response or as a treatment or prophylactic against diseases  -
XX
PS   Disclosure; Page 19; 65pp; English.
XX
CC   The specification describes a peptide variant of lemA, comprising a
CC   hydrophobic element joined to a CD8+ epitope. The peptides may be
CC   used therapeutically by administering the peptides to a patient having
CC   a need to induce a directed CD8+ T cell response. The peptide may also
CC   be used as a preventive measure to avoid a disease or condition, or to
CC   treat subjects already afflicted with a disease. The present sequence
CC   represents a peptide of the invention.
XX
SQ   Sequence   48 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 48;
  Best Local Similarity   100.0%;  Pred. No. 0.78;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db          1 SIINFEKL 8


RESULT 109
AAB48953
ID   AAB48953 standard; Protein; 49 AA.
XX
AC   AAB48953;
XX
DT   27-MAR-2001  (first entry)
XX
DE   Tn5-DICE ovalbumin MHC class I epitope fusion protein.
XX
KW   Transposable element; MHC epitope; major histocompatibility complex;
KW   intracellular bacterial pathogen; loxP site; Cre recombinase;
KW   insertion end; in-frame fusion; detection; antigen;
KW   disseminated insertions of class-I epitopes; DICE-I; transposon Tn5;
KW   ovalbumin MHC class I epitope.
XX
```

```
OS   Synthetic.
XX
PN   WO200071158-A1.
XX
PD   30-NOV-2000.
XX
PF   26-MAY-2000; 2000WO-US14687.
XX
PR   26-MAY-1999;   99US-0136210.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Heffron FL,  Parker DC,  Ellefson DD;
XX
DR   WPI; 2001-031967/04.
XX
PT   Transposable element for detecting an antigenic epitope of a pathogen,
PT   comprising 5' and 3' recombining sites, nucleic acid sequences encoding
PT   a selectable marker and major histocompatibility complex (MHC) epitope,
PT   and an insertion end -
XX
PS   Example 2; Fig 2; 63pp; English.
XX
CC   The invention relates to a novel transposable element comprising DNA
CC   encoding a selectable marker (e.g., antibiotic resistance) located
CC   between a 5' recombining site and a 3' recombining site (e.g., loxP
CC   sites); DNA encoding an MHC (major histocompatibility complex) epitope
CC   either 5' of the 5' recombining site or 3' of the 3' recombining site;
CC   and insertion ends comprising an inverted repeat sequence at the 5' and
CC   3' ends of the transposable element sufficient for integration of the
CC   transposable element. The transposable elements of the invention are able
CC   to introduce in-frame insertions throughout the chromosome of an
CC   intracellular bacterial pathogen. This system "tags" the bacterial gene
CC   and resulting protein, allowing the identification of proteins
CC   secreted across the membranes of the eukaryotic cell infected by the
CC   bacterium. In one embodiment, the transposable elements contain an
CC   antibiotic resistance cassette, two minimal loxP recombination sites, an
CC   MHC class I or class II epitope, and flanking insertion ends. A
CC   transposase, such as the Cre recombinase protein, is expressed in trans
CC   from a plasmid, or can be included in the transposable element. The Cre
CC   recombinase loops out the intervening sequences containing the antibiotic
CC   resistance cassette. When the transposable element inserts within a gene,
CC   the resolved insertion places the MHC class I or class II epitope in
CC   frame with the gene. The transposable elements of the invention are
CC   useful for detecting an antigenic epitope of an intracellular bacterial
CC   pathogen, such as Salmonella sp., Mycobacterium tuberculosis and Listeria
CC   monocytogenes. Certain embodiments of the technology, termed
CC   "disseminated insertions of class-I epitopes" (DICE-I; DICE-II for
CC   class II epitopes) allow the rapid and accurate identification of
CC   proteins involved in bacterial pathogenesis so that such proteins can
CC   be used as vaccine and drug targets. Carrier vaccines may be generated
CC   by infecting bacteria with a transposable element of the invention
CC   which additionally comprises an antigen associated with a disease,
CC   preferably cancer or a viral or bacterial disease, operably linked to the
CC   MHC epitope DNA of the transposable element. The present sequence
CC   represents ovalbumin MHC class I epitope-containing fusion protein
CC   encoded by a Tn5-DICE transposable element.
XX
SQ   Sequence   49 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 0.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         12 SIINFEKL 19

RESULT 110
AAB48954
ID   AAB48954 standard; Protein; 57 AA.
XX
AC   AAB48954;
XX
DT   27-MAR-2001  (first entry)
XX
DE   DICE-I ovalbumin MHC class I epitope-containing fusion protein.
XX
KW   Transposable element; MHC epitope; major histocompatibility complex;
KW   intracellular bacterial pathogen; loxP site; Cre recombinase;
KW   insertion end; in-frame fusion; detection; antigen;
KW   disseminated insertions of class-I epitopes; DICE-I; transposon Tn5;
KW   ovalbumin MHC class I epitope.
XX
OS   Synthetic.
XX
PN   WO200071158-A1.
XX
PD   30-NOV-2000.
XX
PF   26-MAY-2000; 2000WO-US14687.
XX
PR   26-MAY-1999;   99US-0136210.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Heffron FL,  Parker DC,  Ellefson DD;
XX
DR   WPI; 2001-031967/04.
DR   N-PSDB; AAC91698.
XX
PT   Transposable element for detecting an antigenic epitope of a pathogen,
PT   comprising 5' and 3' recombining sites, nucleic acid sequences encoding
PT   a selectable marker and major histocompatibility complex (MHC) epitope,
PT   and an insertion end -
XX
PS   Example 2; Fig 12; 63pp; English.
XX
CC   The invention relates to a novel transposable element comprising DNA
CC   encoding a selectable marker (e.g., antibiotic resistance) located
CC   between a 5' recombining site and a 3' recombining site (e.g., loxP
CC   sites); DNA encoding an MHC (major histocompatibility complex) epitope
CC   either 5' of the 5' recombining site or 3' of the 3' recombining site;
CC   and insertion ends comprising an inverted repeat sequence at the 5' and
CC   3' ends of the transposable element sufficient for integration of the
CC   transposable element. The transposable elements of the invention are able
CC   to introduce in-frame insertions throughout the chromosome of an
CC   intracellular bacterial pathogen. This system "tags" the bacterial gene
CC   and resulting protein, allowing the identification of proteins
CC   secreted across the membranes of the eukaryotic cell infected by the
CC   bacterium. In one embodiment, the transposable elements contain an
CC   antibiotic resistance cassette, two minimal loxP recombination sites, an
CC   MHC class I or class II epitope, and flanking insertion ends. A
CC   transposase, such as the Cre recombinase protein, is expressed in trans
CC   from a plasmid, or can be included in the transposable element. The Cre
CC   recombinase loops out the intervening sequences containing the antibiotic
CC   resistance cassette. When the transposable element inserts within a gene,
CC   the resolved insertion places the MHC class I or class II epitope in
CC   frame with the gene. The transposable elements of the invention are
CC   useful for detecting an antigenic epitope of an intracellular bacterial
CC   pathogen, such as Salmonella sp., Mycobacterium tuberculosis and Listeria
CC   monocytogenes. Certain embodiments of the technology, termed
CC   "disseminated insertions of class-I epitopes" (DICE-I; DICE-II for
CC   class II epitopes) allow the rapid and accurate identification of
CC   proteins involved in bacterial pathogenesis so that such proteins can
CC   be used as vaccine and drug targets. Carrier vaccines may be generated
CC   by infecting bacteria with a transposable element of the invention
CC   which additionally comprises an antigen associated with a disease,
CC   preferably cancer or a viral or bacterial disease, operably linked to the
CC   MHC epitope DNA of the transposable element. The present sequence
CC   represents an ovalbumin MHC class I epitope-containing fusion protein
CC   encoded by the resolved sequence of a DICE-I transposable element.
XX
SQ   Sequence   57 AA;
```

```
  Query Match             100.0%;  Score 38;  DB 22;  Length 57;
  Best Local Similarity   100.0%;  Pred. No. 0.93;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         10 SIINFEKL 17


RESULT 111
AAE13458
ID   AAE13458 standard; Protein; 100 AA.
XX
AC   AAE13458;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken ovalbumin derived protein domain #1.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatibility complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "Javelin sequence"
FT   Region          66..73
FT                   /note= "MHC class I epitope"
FT   Region          74..89
FT                   /note= "MHC class II epitope"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 14; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatibility complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken ovalbumin derived protein domain.
XX
SQ   Sequence   100 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         66 SIINFEKL 73


RESULT 112
AAE13460
ID   AAE13460 standard; Protein; 100 AA.
XX
AC   AAE13460;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken ovalbumin derived protein domain #3.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatibility complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Region          58..65
FT                   /note= "MHC class I epitope"
FT   Region          66..81
FT                   /note= "MHC class II epitope"
FT   Region          93..100
FT                   /note= "Javelin sequence"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 14; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatibility complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken ovalbumin derived protein domain.
XX
SQ   Sequence   100 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 100;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         58 SIINFEKL 65


RESULT 113
AAE13459
ID   AAE13459 standard; Protein; 103 AA.
```

```
XX
AC   AAE13459;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken ovalbumin derived protein domain #2.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatibility complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "Javelin sequence"
FT   Region          9..11
FT                   /note= "Linker"
FT   Region          69..76
FT                   /note= "MHC class I epitope"
FT   Region          77..92
FT                   /note= "MHC class II epitope"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 14; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatibility complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken ovalbumin derived protein domain.
XX
SQ   Sequence   103 AA;

  Query Match            100.0%;  Score 38;  DB 23;  Length 103;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         69 SIINFEKL 76


RESULT 114
AAE13461
ID   AAE13461 standard; Protein; 103 AA.
XX
AC   AAE13461;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken ovalbumin derived protein domain #4.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatibility complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Region          58..65
FT                   /note= "MHC class I epitope"
FT   Region          66..81
FT                   /note= "MHC class II epitope"
FT   Region          93..95
FT                   /note= "Linker sequence"
FT   Region          96..103
FT                   /note= "Javelin sequence"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 14; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatibility complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken ovalbumin derived protein domain.
XX
SQ   Sequence   103 AA;

  Query Match            100.0%;  Score 38;  DB 23;  Length 103;
  Best Local Similarity   100.0%;  Pred. No. 1.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         58 SIINFEKL 65


RESULT 115
AAR89966
ID   AAR89966 standard; Protein; 106 AA.
XX
AC   AAR89966;
XX
DT   25-MAR-2003  (updated)
DT   12-SEP-1996  (first entry)
XX
DE   Polytope sequence.
XX
```

```
KW   Polytope; vaccinia; cytotoxic T lymphocyte; CTL; mouse; epitope; vaccine;
KW   major histocompatability complex; pathogen; HLA diversity; avipox virus;
KW   bacterial vector; rhabdovirus vector; ISCOM; influenza nuclear protein;
KW   ovalbumin; cytomegalovirus; adenovirus; sendai virus; P.Berghei; MHC;
KW   circumsporozoite protein; influenza NS1; choriomeningitis virus.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Region          4..12
FT                   /note= "cytotoxic T lymphocyte (CTL) epitope #1, isolated
FT                   from influenza nuclear protein (residues
FT                   366-374)"
FT   Region          13..20
FT                   /note= "CTL epitope #2, isolated from ovalbumin (residues
FT                   257-264)"
FT   Region          21..29
FT                   /note= "CTL epitope #3, isolated from influenza nuclear
FT                   protein (residues 147-155)"
FT   Region          30..37
FT                   /note= "CTL epitope #4, isolated from influenza nuclear
FT                   protein (residues 50-58)"
FT   Region          38..46
FT                   /note= "CTL epitope #5, isolated from murine
FT                   cytomegalovirus pp89 (residues 168-176)"
FT   Region          48..58
FT                   /note= "CTL epitope #6, isolated from adenovirus 5 E1A
FT                   (residues 234-243)"
FT   Region          59..67
FT                   /note= "CTL epitope #7, isolated from sendai virus
FT                   nuclear protein (residues 324-332)"
FT   Region          68..76
FT                   /note= "CTL epitope #8, isolated from P.Berghei
FT                   circumsporozoite protein (residues 249-257)"
FT   Region          77..85
FT                   /note= "CTL epitope #9, isolated from influenza NS1
FT                   (residues 152-160)"
FT   Misc-difference 78
FT                   /note= "encoded by GAC"
FT   Region          86..94
FT                   /note= "CTL epitope #10, isolated from lymphocytic
FT                   choriomeningitis virus nuclear protein (residues
FT                   118-126)"
FT   Region          97..106
FT                   /note= "monoclonal antibody epitope"
FT   Misc-difference 105
FT                   /note= "encoded by AGA"
XX
PN   WO9603144-A1.
XX
PD   08-FEB-1996.
XX
PF   27-JUL-1995;   95WO-AU00461.
XX
PR   08-FEB-1995;   95AU-0001009.
PR   27-JUL-1994;   94AU-0007079.
XX
PA   (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (UYME ) UNIV MELBOURNE.
PA   (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
PA   (BIOT-) BIOTECH AUSTRALIA PTY LTD.
PA   (CSLC-) CSL LTD.
XX
PI   Burrows SR,  Coupar BEH,  Khanna R,  Moss DJ,  Suhrbier A;
PI   Thomson SA;
XX
DR   WPI; 1996-116788/12.
DR   N-PSDB; AAT12413.
XX
PT   New poly:epitope cytotoxic T lymphocyte vaccines - comprising a
PT   recombinant protein including CTL epitope(s) from pathogens, free of
PT   natural flanking sequences
XX
PS   Claim 5; Fig 5; 46pp; English.
XX
CC   This sequence represents a polytope encoded by a DNA insert of a
CC   recombinant vaccinia virus of the invention.  This sequence contains 10
CC   murine cytotoxic T lymphocyte (CTL) epitopes.  Each of the epitopes in
CC   this sequence are capable of producing a primary CTL response in mice
CC   with the appropriate major histocompatability complex (MHC) allele.  This
CC   sequence (and the DNA encoding it) can be used in vaccines against
CC   multiple epitopes derived from several different pathogens.  The vaccine
CC   could alternatively contain a large number of epitopes from one pathogen
CC   so that HLA diversity of the target population is covered.  The vaccines
CC   can be delivered by vaccinia virus, avipox virus, bacterial or
CC   rhabdovirus vectors, or by virus-like particles.  The proteins are
CC   preferably administered with ISCOMs when they are delivered directly.
CC   The advantage with these vaccines is that they provide a more diverse
CC   immune response.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence   106 AA;

  Query Match             100.0%;  Score 38;  DB 17;  Length 106;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         13 SIINFEKL 20


RESULT 116
AAE13462
ID   AAE13462 standard; Protein; 108 AA.
XX
AC   AAE13462;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken ovalbumin derived protein domain #5.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatibility complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "Javelin sequence"
FT   Region          66..73
FT                   /note= "MHC class I epitope"
FT   Region          74..89
FT                   /note= "MHC class II epitope"
FT   Region          101..108
FT                   /note= "Javelin sequence"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
```

```
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 15; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatibility complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken ovalbumin derived protein domain.
XX
SQ   Sequence   108 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 1.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         66 SIINFEKL 73


RESULT 117
AAE13463
ID   AAE13463 standard; Protein; 111 AA.
XX
AC   AAE13463;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken ovalbumin derived protein domain #6.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatibility complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Region          1..8
FT                   /note= "Javelin sequence"
FT   Region          9..11
FT                   /note= "Linker"
FT   Region          69..76
FT                   /note= "MHC class I epitope"
FT   Region          77..92
FT                   /note= "MHc class II epitope"
FT   Region          104..111
FT                   /note= "Javelin sequence"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Disclosure; Page 15; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatibility complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken ovalbumin derived protein domain.
XX
SQ   Sequence   111 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 111;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         69 SIINFEKL 76


RESULT 118
AAY52575
ID   AAY52575 standard; Protein; 132 AA.
XX
AC   AAY52575;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Amino acid sequence of AOS minigene insert of expression vector pMIN.0.
XX
KW   Chimeric; pan DR epitope; expression vector;
KW   promoter; major histocompatibility complex; MHC; targetting; peptide;
KW   epitope; antigen; presentation; class I; cytosolic pathway;
KW   endoplasmic reticulum; class II; extracellular antigen;
KW   endocytic pathway; helper T lymphocyte; HTL; universal epitope;
KW   cytotoxic T lymphocyte; CTL; immune response; immunogenicity; assay;
KW   vaccine; immunity; infection; pathogen; virus; HIV; HBV; HCV;
KW   hepatitis B; hepatitis C; bacterium; protozoan; tumour cell;
KW   autoimmune disease; activation; antiviral; antimalarial;
KW   immunoprotective; minigene.
XX
OS   Synthetic.
XX
PN   WO9958658-A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999;   99WO-US10646.
XX
PR   13-MAY-1998;   98US-0078904.
PR   15-MAY-1998;   98US-0085751.
XX
PA   (EPIM-) EPIMMUNE INC.
XX
PI   Fikes JD,  Hermanson GG,  Sette A,  Ishioka GY,  Livingston B;
PI   Chesnut RW;
XX
DR   WPI; 2000-039103/03.
DR   N-PSDB; AAZ38634.
XX
PT   Expression vectors encoding major histocompatibility targeting
PT   sequence, used as, e.g. tumor vaccines -
XX
PS   Example 1; Fig 20; 130pp; English.
XX
CC   This sequence represents the amino acid sequence of the AOS minigene
CC   insert of the expression vector pMIN.0 (AAZ38634). This insert encodes
CC   several MHC class I epitopes, and also the universal MHC
CC   class II (helper T) epitope, pan DR epitope (PADRE), and was used
```

```
CC   in an exemplification of the present invention. The invention
CC   relates to a novel expression vector comprising a promoter operably
CC   linked to a fusion gene encoding a major histocompatibility complex
CC   (MHC) targetting sequence, and two or more heterologous peptide
CC   epitopes. The MHC targetting sequence may be a class I targetting
CC   sequence, which directs an MHC class I epitope to a cytosolic pathway or
CC   to the endoplasmic reticulum, or an MHC class II targetting sequence,
CC   which directs extracellular antigens to enter the endocytic pathway to
CC   be processed into antigen peptides for presentation on MHC class II
CC   molecules. The heterologous epitopes may comprise either helper T
CC   lymphocyte (HTL) epitopes, or a cytotoxic T lymphocyte (CTL) epitope and
CC   a universal HTL epitope such as a pan DR epitope (PADRE). The vectors
CC   are useful for stimulating an immune response in vivo, as well as for use
CC   in assaying the human immunogenicity of a human T cell peptide epitope
CC   in vivo in a non-human mammal. They provide a nucleic acid vaccine for
CC   enhancing immunity against infectious pathogens, such as viruses (e.g.,
CC   HIV, hepatitis B (HBV and hepatitis C (HCV)) bacteria, protozoa (e.g.,
CC   Plasmodium falciparum, the cause of malaria) and also tumour cells and
CC   autoimmune diseases. Universal MHC class II epitopes are advantageously
CC   combined with other MHC class I and class II epitopes to increase the
CC   number of cells that are activated in response to a given antigen and
CC   provide a broader population coverage of MHC-reactive alleles.
XX
SQ   Sequence   132 AA;

  Query Match             100.0%;  Score 38;  DB 21;  Length 132;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db         75 SIINFEKL 82


RESULT 119
AAE13435
ID   AAE13435 standard; Protein; 386 AA.
XX
AC   AAE13435;
XX
DT   12-FEB-2002  (first entry)
XX
DE   Chicken ovalbumin containing plurality of epitopes.
XX
KW   Antigenic complex; epitope; heat shock protein; HSP; tether; javelin;
KW   major histocompatibility complex; MHC; therapy; immune response;
KW   malignancy; chicken.
XX
OS   Gallus gallus.
XX
FH   Key             Location/Qualifiers
FT   Domain          200..291
FT   Region          258..265
FT                   /note= "MHC class I epitope"
FT   Region          266..281
FT                   /note= "MHC class II epitope"
XX
PN   WO200179259-A1.
XX
PD   25-OCT-2001.
XX
PF   17-APR-2001; 2001WO-US12567.
XX
PR   17-APR-2000; 2000US-197462P.
XX
PA   (ROTH/) ROTHMAN J E.
PA   (MAYH/) MAYHEW M.
PA   (HOEM/) HOE M.
XX
PI   Rothman JE,  Mayhew M,  Hoe M;
XX
DR   WPI; 2002-017594/02.
DR   N-PSDB; AAD22407.
XX
PT   A new antigenic complex comprising epitopes non-covalently joined to a
PT   heat shock protein by a molecular tether designated a javelin are
PT   useful to treat or prevent infectious disease or malignancy  -
XX
PS   Example; Fig 2; 47pp; English.
XX
CC   The present invention relates to an antigenic complex, comprising a
CC   number of epitopes non-covalently joined to a heat shock protein (HSP) by
CC   a tethering molecule referred to as javelin which has affinity for the
CC   HSP under physiological conditions, where the epitopes are covalently
CC   joined to the tethering molecule and one epitope is major
CC   histocompatibility complex class I (MHC) and the other MHC class II. The
CC   antigenic complex is used to induce immune responses directed towards the
CC   treatment or prevention of infectious diseases and malignancies. The
CC   present sequence is chicken ovalbumin containing plurality of epitopes.
XX
SQ   Sequence   386 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 386;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db        258 SIINFEKL 265


RESULT 120
AAB31545
ID   AAB31545 standard; Protein; 409 AA.
XX
AC   AAB31545;
XX
DT   30-APR-2001  (first entry)
XX
DE   Amino acid sequence of chicken ovalbumin.
XX
KW   Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;
KW   lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;
KW   ovalbumin.
XX
OS   Gallus sp.
XX
PN   WO200104344-A2.
XX
PD   18-JAN-2001.
XX
PF   10-JUL-2000; 2000WO-US18828.
XX
PR   08-JUL-1999;   99US-0143757.
XX
PA   (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI   Siegel M,  Chu NR,  Mizzen LA;
XX
DR   WPI; 2001-138361/14.
DR   N-PSDB; AAF25127.
XX
PT   Screening for compounds that stimulate Th1-like responses in CD4+ T
PT   lymphocyte cells  -
XX
PS   Example 8; Fig 6; 88pp; English.
XX
CC   The present sequence represents an ovalbumin protein. Ovalbumin was fused
CC   to a heat shock protein (Hsp), and used used in the method of the
CC   invention. The specification describes a method of determining whether
CC   a compound stimulates a Th1-like response. Th1 cells are a subset of
CC   CD4+ T lymphocyte cells. The method comprises contacting naive
CC   lymphocytes in vitro with a fusion protein comprising at least a
CC   fragment of Hsp, and then detecting the TH1-like response exhibited by
CC   the cell sample. The proteins which may be used in the method of the
CC   invention are Hsp65, Hsp40, Hsp10, Hsp60, and Hsp71. The method may be
```

```
CC   used to identify compounds that stimulate Th1-like responses in response
CC   to microbial pathogens.
XX
SQ   Sequence   409 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 7.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db        281 SIINFEKL 288


RESULT 121
AAE13112
ID   AAE13112 standard; Protein; 479 AA.
XX
AC   AAE13112;
XX
DT   28-JAN-2002  (first entry)
XX
DE   Human HER300-rGM-CSF fusion construct comprising OVA-derived peptide.
XX
KW   Immunostimulatory fusion protein; IFP; antigen component; therapy;
KW   immunostimulatory component; T-cell mediated immune response; DC;
KW   dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
KW   PAP protein; Ala Arg linker; membrane distal extracellular domain;
KW   membrane distal intracellular domain; C-terminal tag; human; GM-CSF;
KW   HER-2 protein; granulocyte-macrophage colony stimulating factor;
KW   ovalbumin-derived octapeptide; OVA; rat; HER300-rGM-CSF fusion protein.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - Rattus norvegicus.
OS   Chimeric - Unidentified.
XX
PN   WO200174855-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US10515.
XX
PR   30-MAR-2000; 2000US-193504P.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Vidovic D,  Graddis T;
XX
DR   WPI; 2001-662965/76.
DR   N-PSDB; AAD21568.
XX
PT   An immunostimulatory fusion protein comprising the intracellular domain
PT   of HER-2 and an antigen elicits an immune response to the antigen and
PT   is useful for the treatment of associated cancer associated -
XX
PS   Example 1; Page 27; 59pp; English.
XX
CC   The invention relates to immunostimulatory fusion proteins (IFP) and
CC   nucleic acid molecules encoding such proteins. The IFPs comprise a
CC   polypeptide antigen component and an immunostimulatory component derived
CC   from the intracellular domain of HER-2 protein which is effective to
CC   elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
CC   immune response to the antigen. IFP or superactivated dendritic cells
CC   are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
CC   associated with a particularly antigen. The present sequence is HER300
CC   rGM-CSF fusion protein construct which comprises human PAP
CC   signal sequence, mature PAP protein, an Ala Arg linker, human HER-2
CC   signal sequence, mature HER-2 membrane distal extracellular domain,
CC   an Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
CC   an Ala linker, a HER-2 membrane distal intracellular domain, a mature
CC   rat granulocyte-macrophage colony stimulating factor (GM-CSF) sequence
CC   and a C-terminal tag.
XX
SQ   Sequence   479 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 8.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db        331 SIINFEKL 338


RESULT 122
AAU99725
ID   AAU99725 standard; Protein; 541 AA.
XX
AC   AAU99725;
XX
DT   07-OCT-2002  (first entry)
XX
DE   Yeast/mouse SS-OVA-Kb/beta2m-c-myc-AGA2 fusion protein.
XX
KW   Mutant major histocompatibility complex class I chimeric protein; MHC;
KW   lymphocyte; T-cell receptor; tissue sample; biopsy material; pathogen;
KW   bodily fluid; T lymphocyte; neoplastic cell; tumour cell; MHC antigen;
KW   virus; protozoan; bacteria; fungi; nematode; immune response; activator;
KW   enhancer; T cell activator; mouse; recombinant yeast cell; Kb; OVA;
KW   beta2m; dEV8; AGA2; SIYK; fusion protein.
XX
OS   Chimeric - Mus sp.
OS   Chimeric - Saccharomyces cerevisiae.
OS   Synthetic.
XX
PN   WO200246399-A2.
XX
PD   13-JUN-2002.
XX
PF   10-DEC-2001; 2001WO-US47817.
XX
PR   08-DEC-2000; 2000US-254495P.
XX
PA   (UNII ) UNIV ILLINOIS FOUND.
XX
PI   Kranz DM,  Brophy S;
XX
DR   WPI; 2002-527916/56.
DR   N-PSDB; ABK87870.
XX
PT   New isolated mutant major histocompatibility complex class I chimeric
PT   protein displayed on surfaces of recombinant yeast cells, has improved
PT   stability, and is useful for activating immune response  -
XX
PS   Example 7; Page 38-39; 96pp; English.
XX
CC   The present invention relates to a new mutant major histocompatibility
CC   complex (MHC) class I chimeric protein. The protein of the invention
CC   comprises a portion mediating binding to surfaces of recombinant yeast
CC   cells and a portion comprising peptide binding region of MHC class I
CC   protein, where the invention is improved in stability as compared with
CC   MHC class I chimeric protein which is not a mutant chimeric protein.
CC   The protein, further comprising a detectable label, is useful for
CC   detecting a lymphocyte having a T-cell receptor protein in a biological
CC   sample such as cells, tissue sample, biopsy material or bodily fluids.
CC   The method is useful for detecting a T lymphocyte that is specific for
CC   a neoplastic cell, a tumour cell, a virus-infected cell, a protozoan-
CC   infected cell, a bacterium-infected cell or a fungus-infected cell. The
CC   protein of the invention can be used to directly activate T cells, in
CC   order to identify/screen for peptide-MHC antigens. The protein is also
CC   useful in activating T cells that participate in the removal of target
CC   cells including neoplastic cells and cells infected with pathogenic
CC   agents including viruses, protozoans, bacteria, fungi or nematodes.
CC   The invention is improved in stability as compared with MHC class I
CC   protein which is not a mutant chimeric protein. The present amino acid
CC   sequence represents a chimeric MHC protein that is encoded by a
```

```
CC   yeast/mouse fusion gene, as described above.
XX
SQ   Sequence   541 AA;

  Query Match             100.0%;  Score 38;  DB 23;  Length 541;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                ||||||||
Db         34 SIINFEKL 41


RESULT 123
AAE13110
ID   AAE13110 standard; Protein; 564 AA.
XX
AC   AAE13110;
XX
DT   28-JAN-2002  (first entry)
XX
DE   Human HER500 fusion protein construct comprising OVA-derived octapeptide.
XX
KW   Immunostimulatory fusion protein; IFP; antigen component; therapy;
KW   immunostimulatory component; T-cell mediated immune response; DC;
KW   dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
KW   PAP protein; Ala Arg linker; membrane distal extracellular domain;
KW   membrane distal intracellular domain; C-terminal tag; human; OVA;
KW   HER-2 protein; ovalbumin-derived octapeptide; HER500 fusion protein.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - Unidentified.
XX
PN   WO200174855-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US10515.
XX
PR   30-MAR-2000; 2000US-193504P.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Vidovic D,  Graddis T;
XX
DR   WPI; 2001-662965/76.
DR   N-PSDB; AAD21566.
XX
PT   An immunostimulatory fusion protein comprising the intracellular domain
PT   of HER-2 and an antigen elicits an immune response to the antigen and
PT   is useful for the treatment of associated cancer associated -
XX
PS   Claim 7; Page 26; 59pp; English.
XX
CC   The invention relates to immunostimulatory fusion proteins (IFP) and
CC   nucleic acid molecules encoding such proteins. The IFPs comprise a
CC   polypeptide antigen component and an immunostimulatory component derived
CC   from the intracellular domain of HER-2 protein which is effective to
CC   elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
CC   immune response to the antigen. IFP or superactivated dendritic cells
CC   are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
CC   associated with a particularly antigen. The present sequence is HER500
CC   fusion protein construct which comprises human PAP signal
CC   sequence, mature PAP protein, an Ala Arg linker, human HER-2 signal
CC   sequence, mature HER-2 membrane distal extracellular domain, an
CC   Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
CC   HER-2 membrane distal intracellular domain and a C-terminal tag.
XX
SQ   Sequence   564 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 564;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                ||||||||
Db        331 SIINFEKL 338


RESULT 124
AAE13111
ID   AAE13111 standard; Protein; 697 AA.
XX
AC   AAE13111;
XX
DT   28-JAN-2002  (first entry)
XX
DE   Human HER500-rGM-CSF fusion construct comprising OVA-derived peptide.
XX
KW   Immunostimulatory fusion protein; IFP; antigen component; therapy;
KW   immunostimulatory component; T-cell mediated immune response; DC;
KW   dendritic cell; colon cancer; breast carcinoma; ovarian cancer;
KW   PAP protein; Ala Arg linker; membrane distal extracellular domain;
KW   membrane distal intracellular domain; C-terminal tag; human; GM-CSF;
KW   HER-2 protein; granulocyte-macrophage colony stimulating factor;
KW   ovalbumin-derived octapeptide; OVA; rat; HER500-rGM-CSF fusion protein.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - Rattus norvegicus.
OS   Chimeric - Unidentified.
XX
PN   WO200174855-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US10515.
XX
PR   30-MAR-2000; 2000US-193504P.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Vidovic D,  Graddis T;
XX
DR   WPI; 2001-662965/76.
DR   N-PSDB; AAD21567.
XX
PT   An immunostimulatory fusion protein comprising the intracellular domain
PT   of HER-2 and an antigen elicits an immune response to the antigen and
PT   is useful for the treatment of associated cancer associated -
XX
PS   Claim 7; Page 27; 59pp; English.
XX
CC   The invention relates to immunostimulatory fusion proteins (IFP) and
CC   nucleic acid molecules encoding such proteins. The IFPs comprise a
CC   polypeptide antigen component and an immunostimulatory component derived
CC   from the intracellular domain of HER-2 protein which is effective to
CC   elicit a protective dendritic cell (DC)-induced T-cell mediated cellular
CC   immune response to the antigen. IFP or superactivated dendritic cells
CC   are used to treat cancer e.g. breast carcinoma, ovarian and colon cancer
CC   associated with a particularly antigen. The present sequence is HER500
CC   rGM-CSF fusion protein construct which comprises human PAP
CC   signal sequence, mature PAP protein, an Ala Arg linker, human HER-2
CC   signal sequence, mature HER-2 membrane distal extracellular domain,
CC   an Ala linker, an ovalbumin (OVA)-derived immunodominant octapeptide,
CC   HER-2 membrane distal intracellular domain, an Ala Ala linker, a mature
CC   rat granulocyte-macrophage colony stimulating factor (GM-CSF) sequence
CC   and a C-terminal tag.
XX
SQ   Sequence   697 AA;

  Query Match             100.0%;  Score 38;  DB 22;  Length 697;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
                ||||||||
```

```
Db        331 SIINFEKL 338


RESULT 125
AAB31611
ID   AAB31611 standard; Protein; 948 AA.
XX
AC   AAB31611;
XX
DT   30-APR-2001  (first entry)
XX
DE   Amino acid sequence of Hsp65-ovalbumin fusion protein.
XX
KW   Heat shock protein; Hsp; Th1 response; Th1 cell; CD4+ T lymphocyte cell;
KW   lymphocyte; Hsp65; Hsp40; Hsp10; Hsp60; Hsp71; microbial pathogen;
KW   ovalbumin.
XX
OS   Synthetic.
OS   Mycobacterium bovis.
OS   Gallus sp.
XX
PN   WO200104344-A2.
XX
PD   18-JAN-2001.
XX
PF   10-JUL-2000; 2000WO-US18828.
XX
PR   08-JUL-1999;   99US-0143757.
XX
PA   (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.
XX
PI   Siegel M,  Chu NR,  Mizzen LA;
XX
DR   WPI; 2001-138361/14.
DR   N-PSDB; AAF25014.
XX
PT   Screening for compounds that stimulate Th1-like responses in CD4+ T
PT   lymphocyte cells  -
XX
PS   Example 8; Fig 7A-B; 88pp; English.
XX
CC   The present sequence represents a fusion protein comprising a
CC   Mycobacterium bovis heat shock protein (Hsp) 65 fused at its C terminal
CC   to an ovalbumin protein. The fusion protein is used in the method of the
CC   invention. The specification describes a method of determining whether a
CC   compound stimulates a Th1-like response. Th1 cells are a subset of CD4+
CC   T lymphocyte cells. The method comprises contacting naive lymphocytes
CC   in vitro with a fusion protein comprising at least a fragment of Hsp,
CC   and then detecting the TH1-like response exhibited by the cell sample.
CC   The proteins which may be used in the method of the invention are Hsp65,
CC   Hsp40, Hsp10, Hsp60, and Hsp71. The method may be used to identify
CC   compounds that stimulate Th1-like responses in response to microbial
CC   pathogens.
XX
SQ   Sequence   948 AA;

  Query Match            100.0%;  Score 38;  DB 22;  Length 948;
  Best Local Similarity  100.0%;  Pred. No. 17;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 SIINFEKL 8
              ||||||||
Db        820 SIINFEKL 827


Search completed: January 30, 2004, 07:17:23
Job time : 74 secs
```

```
! FINDPATTERNS on geneseqp:* allowing 0 mismatches
!       1 C(R,K){20,20}                               January 30, 2004 07:26 ..

             AAW45801   ck: 9500   len: 39    ! Aaw45801 One chain of a bombesin dimer. 6/1

1                       C(R,K){20,20}
                          C(K){20}
              1:        CKKKKKKKKKKKKKKKKKKKK XGGGQ

             AAB13780   ck: 7317   len: 21    ! Aab13780 Soluble peptide antigen pK. 11/200

1                       C(R,K){20,20}
                          C(K){20}
              1:        CKKKKKKKKKKKKKKKKKKKK

             ABG92659   ck: 5509   len: 58    ! Abg92659 Human DNA-binding protein #85. 11/

1                       C(R,K){20,20}
                          C(K){20}
             35: KFYFV CKKKKKKKKKKKKKKKKKKKK KKK

             AAU18238   ck: 5509   len: 58    ! Aau18238 Novel human DNA-binding protein #8

1                       C(R,K){20,20}
                          C(K){20}
             35: KFYFV CKKKKKKKKKKKKKKKKKKKK KKK

             AAO03766   ck: 8808   len: 81    ! Aao03766 Human polypeptide SEQ ID NO 17658.

1                       C(R,K){20,20}
                          C(K){20}
             48: LTTTA CKKKKKKKKKKKKKKKKKKKK KKKKK

             AAO11210   ck: 863    len: 70    ! Aao11210 Human polypeptide SEQ ID NO 25102.

1                       C(R,K){20,20}
                          C(K){20}
             30: IDLCL CKKKKKKKKKKKKKKKKKKKK KKKKK

 Databases searched:
        Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

     Total finds:            6
    Total length:  158,726,570
 Total sequences:    1,107,863
        CPU time:     04:40.14
```

```
!!SEQUENCE_LIST 1.0
! FINDPATTERNS on geneseqp:* allowing 0 mismatches

!       1 C(R,K){20,20}                           January 30, 2004 07:38 ..

GENESEQP1990S:AAW45801  ck: 9500  len: 39    finds: 1    ! Aaw45801 One chain of a bomb

GENESEQP2000S:AAB13780  ck: 7317  len: 21    finds: 1    ! Aab13780 Soluble peptide ant

GENESEQP2002S:ABG92659  ck: 5509  len: 58    finds: 1    ! Abg92659 Human DNA-binding p

GENESEQP2001S:AAU18238  ck: 5509  len: 58    finds: 1    ! Aau18238 Novel human DNA-bin

GENESEQP2001S:AAO03766  ck: 8808  len: 81    finds: 1    ! Aao03766 Human polypeptide S

GENESEQP2001S:AAO11210  ck: 863   len: 70    finds: 1    ! Aao11210 Human polypeptide S

\\End of list

 Databases searched:
       Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

    Total finds:            6
   Total length:  158,726,570
 Total sequences:    1,107,863
        CPU time:     07:02.46
```

```
!!AA_SEQUENCE 1.0
ID   AAW45801 standard; peptide; 39 AA.
XX
AC   AAW45801;
XX
DT   25-JUN-1998  (first entry)
XX
DE   One chain of a bombesin dimer.
XX
KW   Alpha-melanocyte stimulating hormone; alpha-MSH; receptor agonist;
KW   alpha-MSH-ANT; bombesin; dimer; bivalent agonist; disulphide bond;
KW   G-protein coupled receptor.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Cross-links     1
FT                   /note= "This residue is disulphide bonded to the
FT                          corresponding Cys residue of an identical chain
FT                          to form a dimer"
FT   Modified-site   28
FT                   /note= "Epsilon-aminohexanoic acid"
XX
PN   WO9803632-A1.
XX
PD   29-JAN-1998.
XX
PF   23-JUL-1997;   97WO-US12911.
XX
PR   24-JUL-1996;   96US-0686934.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Carrithers MD,  Lerner MR;
XX
DR   WPI; 1998-120757/11.
XX
PT   Bivalent agonist of G-protein coupled receptors containing two
PT   ligand domains - bonded to molecular backbone, for treatment of
PT   hypotension, promotion of skin tanning etc., also for delivering
PT   drugs and gene therapy vectors to selected cells
XX
PS   Claim 41; Page 48; 71pp; English.
XX
CC   This sequence represents one of two identical chains disulphide bonded
CC   to form a bombesin dimer. The invention relates to bivalent agonists,
CC   with affinity for at least 1 G-protein coupled receptor (GPCR). The
CC   bivalent agonists comprise: (a) two ligand domains (LD), individually
CC   agonists or antagonists for GPCR, spaced 40-250 Angstrom apart, and (b)
CC   a molecular backbone (MB) covalently bound to LD. The bivalent agonists
CC   are useful in human or veterinary medicine as carriers for drugs or gene
CC   therapy vectors, allowing these to be endocytosed by GPCR-expressing
CC   cells. They can also be used e.g. to treat hypertension (angiotensin-
CC   based LD); to increase levels of luteinising hormone (LH), using LD
CC   derived from LH-releasing hormone, or to promote skin tanning (LD based
CC   on alpha -melanocyte-stimulating hormone, MSH). The bivalent agonists
CC   are administered orally, by injection or topically. Typical doses for
CC   skin tanning are 1-4000 (especially 30-100) mu mole/kg systemically or
CC   the bivalent agonists are administered topically in a composition
CC   containing 0.001-10 (especially 1) mM. Where both LD are agonists, the
CC   bivalent agonist has a synergistically higher activity than two
CC   individual agonist ligands, and where at least 1 is an antagonist the
CC   effect is stimulatory. The bivalent agonists are active at lower
CC   concentrations than known agonists so should avoid toxicity problems.
XX
SQ   Sequence   39 AA;

 AAW45801  Length: 39  January 30, 2004 07:48  Type: P  Check: 9500  ..

       1  CKKKKKKKKK KKKKKKKKKK KXGGGQQRLG NQWAVGHLM

!!AA_SEQUENCE 1.0
ID   AAB13780 standard; peptide; 21 AA.
XX
AC   AAB13780;
XX
DT   10-NOV-2000  (first entry)
XX
DE   Soluble peptide antigen pK.
XX
KW   pk peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW   major histocompatibility complex class 1; MHC class 1; antigen; tumour;
KW   prostate; breast; multiple myeloma.
XX
OS   Unidentified.
XX
PN   WO200035949-A1.
XX
PD   22-JUN-2000.
XX
PF   14-DEC-1999;   99WO-US29724.
XX
PR   14-DEC-1998;   98US-0112324.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Hakim I,  Vidovic D;
XX
DR   WPI; 2000-442365/38.
XX
PT   Antigens modified by the covalent addition of a peptide that
PT   facilitates entry into antigen presenting cells, useful for producing
PT   compositions for immunizing against tumors and pathogens -
XX
PS   Claim 2; Page 26; 34pp; English.
XX
CC   The present invention relates to compositions of modified soluble protein
CC   antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC   response i.e. a major histocompatability complex (MHC) class I molecule
CC   response. The protein antigen is modified by the covalent addition of a
CC   peptide sequence which facilitate entry of the antigen into antigen
CC   presenting cells (APCs). The present sequence is one such peptide
CC   sequence which can be used to modify the soluble antigens. The present
CC   sequence is peptide pK. The modified antigen composition may be used for
CC   immunising against, or treating a tumour e.g. prostate and breast
CC   carcinoma or multiple myeloma, or pathogen in mammals.
XX
SQ   Sequence   21 AA;

 AAB13780  Length: 21  January 30, 2004 07:48  Type: P  Check: 7317  ..

       1  CKKKKKKKKK KKKKKKKKKK K

!!AA_SEQUENCE 1.0
ID   ABG92659 standard; Protein; 58 AA.
XX
AC   ABG92659;
XX
DT   18-NOV-2002  (first entry)
XX
DE   Human DNA-binding protein #85.
XX
KW   Human; DNA-binding protein; B cell immunodeficiency; autoimmune disorder;
KW   severe combined immunodeficiency; rheumatoid arthritis; Crohn's disease;
KW   diabetes mellitus; allergy; asthma; inflammatory condition; thrombosis;
KW   graft-versus-host disease; blood-related disorder; atherosclerosis;
KW   hyperproliferative disorder; cancer; renal disorder; arrhythmia;
KW   acute glomerulonephritis; cardiovascular disorder; respiratory disorder;
KW   Goodpasture's syndrome; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; endocrine disorder; Addison's disease;
KW   reproductive system disorder; endometriosis; infectious disease;
KW   viral infection; bacterial infection; fungal infection; vaccine;
KW   gastrointestinal disorder; multiple sclerosis; gene therapy.
XX
OS   Homo sapiens.
XX
```

```
PN   US2002102638-A1.
XX
PD   01-AUG-2002.
XX
PF   17-JAN-2001; 2001US-0764846.
XX
PR   31-JAN-2000; 2000US-179065P.
PR   04-FEB-2000; 2000US-180628P.
PR   28-JUN-2000; 2000US-214886P.
PR   07-JUL-2000; 2000US-216647P.
PR   07-JUL-2000; 2000US-216880P.
PR   11-JUL-2000; 2000US-217487P.
PR   11-JUL-2000; 2000US-217496P.
PR   14-JUL-2000; 2000US-218290P.
PR   26-JUL-2000; 2000US-220963P.
PR   26-JUL-2000; 2000US-220964P.
PR   14-AUG-2000; 2000US-224518P.
PR   14-AUG-2000; 2000US-224519P.
PR   14-AUG-2000; 2000US-225267P.
PR   14-AUG-2000; 2000US-225268P.
PR   14-AUG-2000; 2000US-225270P.
PR   14-AUG-2000; 2000US-225447P.
PR   14-AUG-2000; 2000US-225757P.
PR   14-AUG-2000; 2000US-225758P.
PR   22-AUG-2000; 2000US-226868P.
PR   30-AUG-2000; 2000US-228924P.
PR   01-SEP-2000; 2000US-229287P.
PR   01-SEP-2000; 2000US-229343P.
PR   01-SEP-2000; 2000US-229344P.
PR   01-SEP-2000; 2000US-229345P.
PR   05-SEP-2000; 2000US-229509P.
PR   05-SEP-2000; 2000US-229513P.
PR   08-SEP-2000; 2000US-231413P.
PR   21-SEP-2000; 2000US-234223P.
PR   21-SEP-2000; 2000US-234274P.
PR   25-SEP-2000; 2000US-234997P.
PR   27-SEP-2000; 2000US-235834P.
PR   29-SEP-2000; 2000US-236327P.
PR   29-SEP-2000; 2000US-236367P.
PR   29-SEP-2000; 2000US-236368P.
PR   29-SEP-2000; 2000US-236369P.
PR   29-SEP-2000; 2000US-236370P.
PR   02-OCT-2000; 2000US-236802P.
PR   02-OCT-2000; 2000US-237037P.
PR   02-OCT-2000; 2000US-237038P.
PR   02-OCT-2000; 2000US-237039P.
PR   02-OCT-2000; 2000US-237040P.
PR   13-OCT-2000; 2000US-239935P.
PR   20-OCT-2000; 2000US-240960P.
PR   20-OCT-2000; 2000US-241785P.
PR   20-OCT-2000; 2000US-241809P.
PR   01-NOV-2000; 2000US-244617P.
PR   17-NOV-2000; 2000US-249299P.
PR   08-DEC-2000; 2000US-251856P.
PR   08-DEC-2000; 2000US-251868P.
PR   08-DEC-2000; 2000US-251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2002-690611/74.
DR   N-PSDB; ABS68254.
XX
PT   Novel DNA-binding protein useful for diagnosis, prognosis, prevention
PT   and treatment of immune, hyperproliferative, respiratory,
PT   cardiovascular, reproductive, endocrine, gastrointestinal and
PT   neurological disorders  -
XX
PS   Claim 11; SEQ ID No 223; 225pp; English.
XX
CC   The present invention relates to a new DNA-binding protein. The invention
CC   is useful in treating, preventing, diagnosing and/or prognosing
CC   immunodeficiencies (e.g. B cell immunodeficiencies, severe combined
CC   immunodeficiencies), autoimmune disorders (rheumatoid arthritis, multiple
CC   sclerosis, diabetes mellitus), allergic reactions and conditions (e.g.
CC   asthma), inflammatory conditions, graft-versus-host disease, blood-
CC   related disorders (thrombosis, atherosclerosis), hyperproliferative
CC   disorders (e.g. cancer), renal disorders (e.g. acute glomerulonephritis),
CC   cardiovascular disorders (e.g. arrhythmia), respiratory disorders
CC   (Goodpasture's syndrome), neurological disorders (e.g. Alzheimer's
CC   disease, Parkinson's disease), endocrine disorders (e.g. Addison's
CC   disease), reproductive system disorders (e.g. endometriosis),
CC   infectious diseases (e.g. viral, bacterial or fungal infections) and
CC   gastrointestinal disorders (e.g. Crohn's disease). The invention is also
CC   useful to stimulate neuronal growth and treat, prevent, and/or diagnose
CC   neuronal damage which occurs in certain neuronal disorders or neuro-
CC   degenerative conditions. The present amino acid sequence represents a
CC   human DNA-binding protein of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from USPTO at
CC   http.seqdata.uspto.gov/sequence.
XX
SQ   Sequence   58 AA;

 ABG92659  Length: 58  January 30, 2004 07:48  Type: P  Check: 5509  ..

       1  TYLECEHNSL VNSKCLTVVL SRCISVCLNK FYFVCKKKKK KKKKKKKKKK

      51  KKKKKKKK

!!AA_SEQUENCE 1.0
ID   AAU18238 standard; Protein; 58 AA.
XX
AC   AAU18238;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Novel human DNA-binding protein #85.
XX
KW   Human; DNA-binding protein; histone; chromo domain protein;
KW   chromatin organisation modifier; Y-box binding protein;
KW   DNA organisation; gene transcription; malignant disease;
KW   autoimmune disorder; rheumatic disease; genetic abnormality;
KW   infectious disease; neurological disorder; gene therapy;
KW   immunomodulatory; anti-HIV; anti rheumatic; anti microbial;
KW   cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200155162-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01305.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
```

```
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465557/50.
DR   N-PSDB; AAS29114.
XX
PT   Nucleic acid molecules encoding human secreted chromosomal binding
PT   proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT   Alzheimer's and Parkinson's diseases and cancers -
XX
PS   Claim 11; SEQ ID No 223; 561pp; English.
XX
CC   The present invention relates to the isolation of novel DNA-binding
CC   proteins, and cDNA (AAS29030-AAS29157) and genomic sequences encoding
CC   for these proteins. DNA-binding proteins such as histones, chromo
CC   (chromatin organisation modifier) domain proteins, and Y-box binding
CC   proteins may contribute to diseases resulting from aberrant DNA
CC   organisation and/or gene transcription. The sequences of the invention
CC   are useful in screening assays to identify antagonists and/or agonists
CC   that may enhance or block activities mediated by DNA-binding proteins.
CC   Blockers of DNA-binding proteins may be useful in treating disorders
CC   such as malignant diseases (e.g. cancer), autoimmune disorders
CC   (e.g. diabetes mellitus), rheumatic diseases (e.g. rheumatoid
CC   arthritis), genetic abnormalities (e.g. cystic fibrosis), infectious
CC   diseases (e.g. HIV) and neurological disorders (e.g. Alzheimer's
```

```
CC   disease). The polynucleotide sequences of the invention may also be
CC   used in gene therapy. AAU18154-AAU18281 represent novel DNA-binding
CC   proteins.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   58 AA;

 AAU18238  Length: 58  January 30, 2004 07:48  Type: P  Check: 5509  ..

      1  TYLECEHNSL VNSKCLTVVL SRCISVCLNK FYFVCKKKKK KKKKKKKKKK

     51  KKKKKKKK

!!AA_SEQUENCE 1.0
ID   AAO03766 standard; Protein; 81 AA.
XX
AC   AAO03766;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 17658.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI83697.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 17658; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   81 AA;

 AAO03766  Length: 81 .January 30, 2004 07:48  Type: P  Check: 8808  ..

      1  GLNQTQLRKI LAYSSITHIG XIIAVLPYNP NITILNLTIY IILTTTACKK

     51  KKKKKKKKKK KKKKKKKKKK KKKKKKGGGG A

!!AA_SEQUENCE 1.0
ID   AAO11210 standard; Protein; 70 AA.
XX
AC   AAO11210;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 25102.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI91141.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 25102; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   70 AA;

 AAO11210  Length: 70  January 30, 2004 07:48  Type: P  Check: 863  ..

      1  YYIHRTTVFM CMNXGLKDNV DKXTIDLCLC KKKKKKKKKK KKKKKKKKKK

     51  KKKKKKKKKK KKKKPQGGGA
```

```
! FINDPATTERNS on pir:* allowing 0 mismatches

!        1 C(R,K){20,20}                          January 30, 2004 06:58 ..


 Databases searched:
        NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

     Total finds:              0
    Total length:     96,168,682
 Total sequences:        283,308
        CPU time:       01:56.42
```

```
! FINDPATTERNS on swp:* allowing 0 mismatches

!        1 C(R,K){20,20}                          January 30, 2004 06:58 ..


 Databases searched:
        SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
        SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

     Total finds:               0
    Total length:     305,079,309
 Total sequences:         958,388
        CPU time:        06:50.54
```

```
> O <
O| |O  IntelliGenetics
> O <

Quest - Quick User-directed Expression Search Tool
Release 5.4

                 -- Outline of search "seq6-iss" --

Selected search type is key against sequence data banks or files.
Selected scope is Sequence.
Selected sequence key from "new.key":
  seq6 (AA) ID   seq6 AA preliminary pattern
1       followed by
2         c
2         r or k repeated 20 times

Selected data banks and files:

   Data bank : Issued_AA , all entries

                      -- Output Parameters --

Format Options:                       File Options:
  Nucleic acid code matching  Exact     Indirect file             No
  Find non-matching hits only No        Sequence or key file      No
  Report key used             Yes       List of hits              No
  Note position of hit        Yes       Hit display               Yes
  Display full annotations    Yes       Name and annotations      Yes
  Sequence context            50


                        -- Run Parameters --

  Run mode                    Batch
    Time to start comparison  now
    Notify at end of run      No


No hits found.


                      -- Search Statistics --

Times:                  CPU                     Total Elapsed
                    00:01:29.16                 00:01:30.00

Number of sequences searched:                    328807
Number of sequence hits:                              0
Number of separate matches:                           0
Number of sequence hits saved:                        0
```

```
! FINDPATTERNS on pir:* allowing 0 mismatches

!       1 (R,K){20,20}                         January 30, 2004 07:05 ..

        T49173  ck: 4143  len: 517   ! hypothetical protein T20N10.250 - Arabidops

1                  (R,K){20,20}
                    (K){20}
      444: FERVG KKKKKKKKKKKKKKKKKKKK KKKIR

                    (K){20}
      445: ERVGK KKKKKKKKKKKKKKKKKKKK KKIRL

                    (K){20}
      446: RVGKK KKKKKKKKKKKKKKKKKKKK KIRLN

                    (K){20}
      447: VGKKK KKKKKKKKKKKKKKKKKKKK IRLNF

        T46395  ck: 7330  len: 380   ! hypothetical protein DKFZp434I1120.1 - huma

1                  (R,K){20,20}
                    (K){20}
      355: NLLLQ KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
      356: LLLQK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
      357: LLQKK KKKKKKKKKKKKKKKKKKKK KKKK

                    (K){20}
      358: LQKKK KKKKKKKKKKKKKKKKKKKK KKK

                    (K){20}
      359: QKKKK KKKKKKKKKKKKKKKKKKKK KK

                    (K){20}
      360: KKKKK KKKKKKKKKKKKKKKKKKKK K

                    (K){20}
      361: KKKKK KKKKKKKKKKKKKKKKKKKK

        I52523  ck: 8048  len: 215   ! nucleoporin p62 homolog - rat (fragment)

1                  (R,K){20,20}
                    (K){20}
       35: CEFLE KKKKKKKKKKKKKKKKKKKK KKTGD

                    (K){20}
       36: EFLEK KKKKKKKKKKKKKKKKKKKK KTGDN

                    (K){20}
       37: FLEKK KKKKKKKKKKKKKKKKKKKK TGDNA

        S58321  ck: 1384  len: 126   ! probable membrane protein YOR309c - yeast (

1                  (R,K){20,20}
                   (R,K){20}
       53: RKRRT RRKRRKRRKRRRRKRRKRRR KRSPR

                   (R,K){20}
       54: KRRTR RKRRKRRKRRRRKRRKRRRK RSPRK

                   (R,K){20}
       55: RRTRR KRRKRRKRRRRKRRKRRRKR SPRKR

 Databases searched:
```

```
NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

       Total finds:          17
      Total length:  96,168,682
   Total sequences:     283,308
          CPU time:    03:26.44
```

```
!!SEQUENCE_LIST 1.0
! FINDPATTERNS on pir:* allowing 0 mismatches

!        1 (R,K){20,20}                          January 30, 2004 07:51 ..

PIR2:T49173          ck: 4143  len: 517    finds: 4    ! hypothetical protein T20N10.25
PIR2:T46395          ck: 7330  len: 380    finds: 7    ! hypothetical protein DKFZp434I
PIR2:I52523          ck: 8048  len: 215    finds: 3    ! nucleoporin p62 homolog - rat
PIR2:S58321          ck: 1384  len: 126    finds: 3    ! probable membrane protein YOR3
\\End of list

 Databases searched:
        NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

     Total finds:            17
    Total length:     96,168,682
 Total sequences:        283,308
        CPU time:       03:53:65
```

```
!!AA_SEQUENCE 1.0
P1;T49173 - hypothetical protein T20N10.250 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 08-Dec-2000
C;Accession: T49173
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.;
 Rudd, S.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z25017
A;Accession: T49173
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-517 <DAN>
A;Cross-references: EMBL:AL353032; GSPDB:GN00061; ATSP:T20N10.250
A;Experimental source: cultivar Columbia; BAC clone T20N10
C;Genetics:
A;Gene: ATSP:T20N10.250
A;Map position: 3
A;Introns: 312/3; 359/3; 444/3
C;Superfamily: Arabidopsis thaliana hypothetical protein F17J16.30

 T49173  Length: 517  January 30, 2004 07:58  Type: P  Check: 4143  ..

       1  MDLFSSLPNE LLYHILSFLS TKEAALTSVL SKRWRNLFAF VPYLEFDDSV

      51  FLHPEERKRE KEGILQSFMD FVDRVLDLHG DSLIKTFSLK CKTGVDSDHV

     101  DRWICNVLAR GVSDLDLFID FRDLYSLPHE VGVSRTLVVL RVGSESDLYW

     151  WQKFLCLPML KTLVLDSCWL CIGQFQILLL ACPALEELDM TNTRWKDSNV

     201  TVSSSILKEL TIDLHGCCSV VNLKSLSFDA PSLVYFYYCD SLAEDYPQVN

     251  LKNLVEAQIN LLLTQAQIEQ VRALNNEMLV ADDVFPGLGN AWKLITGLRN

     301  VQQLYLSPDT LEVLSRCCEG MPVFNNLKVL SIWSDMNRGW QAMPVLLRNC

     351  PHLETLIIEG LLHYATDKCG DVCDCISRDY KDHSLTSCPV KKLQIYEFRG

     401  TIRELEMIKH FLKIFPCLKE MDIYAHENSH TLFKDPTIFE RVGKKKKKKK

     451  KKKKKKKKKK KKKKKKIRLN FKPVNKTEQF LKRLADKLCF IPQCLEFLDV

     501  DSSLGELALL AMDSRPS

!!AA_SEQUENCE 1.0
F1;T46395 - hypothetical protein DKFZp434I1120.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46395
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46395
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-380 <AAA>
A;Cross-references: EMBL:AL137556
A;Experimental source: adult testis; clone DKFZp434I1120
C;Genetics:
A;Note: DKFZp434I1120.1

 T46395  Length: 380  January 30, 2004 07:58  Type: P  Check: 7330  ..

       1  MGSTDSKLNF RKAVIQLTTK TQPVEATDDA FWDQFWADTA TSVQDVFALV

      51  PAAEIRAVRE ESPSNLATLC YKAVEKLVQG AESGCHSEKE KQIVLNCSRL

     101  LTRVLPYIFE DPDWRGFFWS TVPGAGRGGG EEDDEHARPL AESLLLAIAD

     151  LLFCPDFTVQ SHRRSTVDSA EDVHSLDSCE YIWEAGVGFA HSPQPNYIHD

     201  MNRMELLKLL LTCFSEAMYL PPAPESGSTN PWVQFFCSTE NRHALPLFTS

     251  LLNTVCAYDP VGYGIPYNHL LFSDYREPLV EEAAQVLIVT LDHDSASSAS

     301  PTVDGTTTGT AMDDADPPGP ENLFVNYLSR IHREEDFQFI LKGIARLLSN

     351  LLLQKKKKKK KKKKKKKKKK KKKKKKKKKK

!!AA_SEQUENCE 1.0
F1;I52523 - nucleoporin p62 homolog - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
C;Accession: I52523
R;Wang, Z.Q.; Akmal, K.M.; Kim, K.H.
Biol. Reprod. 51, 1022-1030, 1994
A;Title: An unusual nucleoporin-related messenger ribonucleic acid is present
 in the germ cells of rat testis.
A;Reference number: I52523; MUID:95151924; PMID:7849178
A;Accession: I52523
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-215 <RES>
A;Cross-references: GB:S75997; NID:g913245; PIDN:AAB33384.1; PID:g913246
A;Experimental source: testis

 I52523  Length: 215  January 30, 2004 07:58  Type: P  Check: 8048  ..

       1  SGGRATSSCD EDCLSSSLPF SLSGPVKQDC EFLEKKKKKK KKKKKKKKKK

      51  KKKKKKTGDN AKSVSRQYSL KVTKLEHEAE QAKVELDFIL SQQKELEDLL

     101  SPLEESVKEQ SGTIYLQHAD EEREKTYKLA ENIDAQLKRM AQDLKDIIEH

     151  LNMAGGPADT SDPLQQICKI LNAHMDSLQW VDQSSALLQR RVEEASRVCE

     201  SRRKEQERSL RIAFD

!!AA_SEQUENCE 1.0
P1;S58321 - probable membrane protein YOR309c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein O6105
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002
C;Accession: S58321; S67215; S71989
R;Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.
submitted to the EMBL Data Library, August 1995
A;Reference number: S58318
A;Accession: S58321
A;Molecule type: DNA
A;Residues: 1-126 <PEA>
A;Cross-references: EMBL:X90565; NID:g940836; PID:g940840
R;Pearson, B.M.; Hernando, Y.; Kalogeropoulos, A.; Schweizer, M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67213
A;Accession: S67215
A;Molecule type: DNA
A;Residues: 1-126 <PEW>
A;Cross-references: EMBL:Z75217; NID:g1420680; PID:e252431; PID:g1420681;
 MIPS:YOR309c
A;Experimental source: strain S288C
R;Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.;
 Schweizer, M.
Yeast 12, 1021-1031, 1996
A;Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast
 chromosome XV reveals regions of similarity to chromosomes I and XIII.
A;Reference number: S71986; MUID:97051589; PMID:8896266
A;Accession: S71989
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-126 <PEF>
A;Cross-references: EMBL:X90565; NID:g940836; PIDN:CAA62164.1; PID:g940840
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August
 1995
C;Genetics:
A;Cross-references: SGD:S0005836
```

```
A;Map position: 15R
C;Keywords: transmembrane protein
F;3-19/Domain: transmembrane #status predicted <TM1>
F;107-123/Domain: transmembrane #status predicted <TM2>

 S58321  Length: 126  January 30, 2004 07:58  Type: P  Check: 1384  ..

       1  MQMLIPQRLL LILNPILMMK RKKRKKRKKR RERETMMKIP RILKKLRRKR

      51  RTRRKRRKRR KRRRRKRRKR RRKRSPRKRR KRRNKDAFYI LIISDPSRSL

     101  LFGFRKFSII IQCLTYFSFH ILFHNL
```

```
! FINDPATTERNS on swp:* allowing 0 mismatches
!        1 (R,K){20,20}                          January 30, 2004 07:00 ..

          Q12444   ck: 1384   len: 126    ! Q12444 saccharomyces cerevisiae (baker's ye
1                    (R,K){20,20}
                      (R,K){20}
       53: RKRRT RRKRRKRRKRRRRKRRKRRR KRSPR

                      (R,K){20}
       54: KRRTR RKRRKRRKRRRRKRRKRRRK RSPRK

                      (R,K){20}
       55: RRTRR KRRKRRKRRRRKRRKRRRKR SPRKR

          Q9P529   ck: 291    len: 128    ! Q9p529 neurospora crassa. hypothetical 15.2
1                    (R,K){20,20}
                      (K){20}
       71: KRKNQ KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
       72: RKNQK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
       73: KNQKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
       74: NQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
       75: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
       76: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
       77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
       78: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
       79: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

                      (K){20}
       80: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEQ

                      (K){20}
       81: KKKKK KKKKKKKKKKKKKKKKKKKK KKEQE

                      (K){20}
       82: KKKKK KKKKKKKKKKKKKKKKKKKK KEQES

                      (K){20}
       83: KKKKK KKKKKKKKKKKKKKKKKKKK EQESR

          Q9NT34   ck: 7330   len: 380    ! Q9nt34 homo sapiens (human). hypothetical p
1                    (R,K){20,20}
                      (K){20}
      355: NLLLQ KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
      356: LLLQK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
      357: LLQKK KKKKKKKKKKKKKKKKKKKK KKKK

                      (K){20}
      358: LQKKK KKKKKKKKKKKKKKKKKKKK KKK

                      (K){20}
      359: QKKKK KKKKKKKKKKKKKKKKKKKK KK

                      (K){20}
      360: KKKKK KKKKKKKKKKKKKKKKKKKK K

                      (K){20}
      361: KKKKK KKKKKKKKKKKKKKKKKKKK

          Q9H6Q7   ck: 3351   len: 720    ! Q9h6q7 homo sapiens (human). hypothetical p
1                    (R,K){20,20}
                      (K){20}
      692: IVSIS KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
      693: VSISK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
      694: SISKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
      695: ISKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
      696: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
      697: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                      (K){20}
      698: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                      (K){20}
      699: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                      (K){20}
      700: KKKKK KKKKKKKKKKKKKKKKKKKK K

                      (K){20}
      701: KKKKK KKKKKKKKKKKKKKKKKKKK

          Q8N6F0   ck: 9898   len: 55     ! Q8n6f0 homo sapiens (human). similar to loc
1                    (R,K){20,20}
                      (R,K){20}
       14: RRGRG KKKKRKRKKKKKKKKKKKKK KKKKK

                      (R,K){20}
       15: RGRGK KKKRKRKKKKKKKKKKKKKK KKKKK

                      (R,K){20}
       16: GRGKK KKRKRKKKKKKKKKKKKKKK KKKKK

                      (R,K){20}
       17: RGKKK KRKRKKKKKKKKKKKKKKKK KKKKK

                      (R,K){20}
       18: GKKKK RKRKKKKKKKKKKKKKKKKK KKKKR

                      (R,K){20}
       19: KKKKR KRKKKKKKKKKKKKKKKKKK KKKRR

                      (R,K){20}
       20: KKKRK RKKKKKKKKKKKKKKKKKKK KKRRR

                      (K){20}
       21: KKRKR KKKKKKKKKKKKKKKKKKKK KRRRR
```

```
                (K){20}
 22: KRKRK KKKKKKKKKKKKKKKKKKKK RRRRR

               (R,K){20}
 23: RKRKK KKKKKKKKKKKKKKKKKKKR RRRRR

               (R,K){20}
 24: KRKKK KKKKKKKKKKKKKKKKKKRR RRRRG

               (R,K){20}
 25: RKKKK KKKKKKKKKKKKKKKKKRRR RRRGR

               (R,K){20}
 26: KKKKK KKKKKKKKKKKKKKKKRRRR RRGRR

               (R,K){20}
 27: KKKKK KKKKKKKKKKKKKKKRRRRR RGRRR

               (R,K){20}
 28: KKKKK KKKKKKKKKKKKKKRRRRRR GRRRR

   Q9HC48   ck: 7602   len: 667   ! Q9hc48 homo sapiens (human). ctcl tumor ant

1             (R,K){20,20}
               (R,K){20}
648: GDKTD RKKKKKKKKKKKKKKKKKKK

   Q9H5V6   ck: 379   len: 168   ! Q9h5v6 homo sapiens (human). hypothetical p

1             (R,K){20,20}
                (K){20}
140: VREWE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
141: REWEK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
142: EWEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
143: WEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
144: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
145: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
146: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
147: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
148: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
149: KKKKK KKKKKKKKKKKKKKKKKKKK

   Q8SWR7   ck: 2372   len: 515   ! Q8swr7 drosophila melanogaster (fruit fly).

1             (R,K){20,20}
                (K){20}
493: FTDYI KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
494: TDYIK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
495: DYIKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
496: YIKKK KKKKKKKKKKKKKKKKKKKK

   Q8T2U7   ck: 8768   len: 791   ! Q8t2u7 dictyostelium discoideum (slime mold

1             (R,K){20,20}
                (K){20}
769: EIEKE KKKKKKKKKKKKKKKKKKKK KEI

                (K){20}
770: IEKEK KKKKKKKKKKKKKKKKKKKK EI

   Q8I247   ck: 5951   len: 206   ! Q8i247 plasmodium falciparum (isolate 3d7).

1             (R,K){20,20}
                (K){20}
185: KKIFL KKKKKKKKKKKKKKKKKKKK KT

                (K){20}
186: KIFLK KKKKKKKKKKKKKKKKKKKK T

   Q95LV6   ck: 7515   len: 531   ! Q95lv6 macaca fascicularis (crab eating mac

1             (R,K){20,20}
                (K){20}
502: YKGNS KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
503: KGNSK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
504: GNSKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
505: NSKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
506: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
507: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
508: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
509: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
510: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
511: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
512: KKKKK KKKKKKKKKKKKKKKKKKKK

   Q9LXR2   ck: 4143   len: 517   ! Q9lxr2 arabidopsis thaliana (mouse-ear cres

1             (R,K){20,20}
                (K){20}
444: FERVG KKKKKKKKKKKKKKKKKKKK KKKIR

                (K){20}
445: ERVGK KKKKKKKKKKKKKKKKKKKK KKIRL
```

```
                  (K){20}
       446: RVGKK KKKKKKKKKKKKKKKKKKKK KIRLN

                  (K){20}
       447: VGKKK KKKKKKKKKKKKKKKKKKKK IRLNF

  Q8S7D3  ck: 6479  len: 80   ! Q8s7d3 oryza sativa (rice). hypothetical 9.

1               (R,K){20,20}
                  (K){20}
        48: VIHLD KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        49: IHLDK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        50: HLDKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        51: LDKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        52: DKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKL

                  (K){20}
        54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKLL

                  (K){20}
        55: KKKKK KKKKKKKKKKKKKKKKKKKK KKLLG

                  (K){20}
        56: KKKKK KKKKKKKKKKKKKKKKKKKK KLLGE

                  (K){20}
        57: KKKKK KKKKKKKKKKKKKKKKKKKK LLGE

  Q8LQP6  ck: 3239  len: 113   ! Q8lqp6 oryza sativa (japonica cultivar-grou

1               (R,K){20,20}
                  (K){20}
        10: SLEHI KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        11: LEHIK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        12: EHIKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        13: HIKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                  (K){20}
        14: IKKKK KKKKKKKKKKKKKKKKKKKK KKKRE

                  (K){20}
        15: KKKKK KKKKKKKKKKKKKKKKKKKK KKRER

                  (K){20}
        16: KKKKK KKKKKKKKKKKKKKKKKKKK KRERE

                  (K){20}
        17: KKKKK KKKKKKKKKKKKKKKKKKKK RERER

                (R,K){20}
        18: KKKKK KKKKKKKKKKKKKKKKKKKR ERERE

  Q9LGZ9  ck: 6094  len: 260   ! Q9lgz9 arabidopsis thaliana (mouse-ear cres
```

```
1                 (R,K){20,20}
                    (R,K){20}
         6: MDRCI RKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         7: DRCIR KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         8: RCIRK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         9: CIRKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        10: IRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        11: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
74: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
```

```
 78: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 79: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 80: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 81: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 82: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 87: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 89: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 90: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 91: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 92: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
 99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
108: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
109: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
110: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
111: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
112: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
113: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
115: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
116: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
117: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
118: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
121: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
122: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
123: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
124: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
125: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
126: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
(K){20}
127: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
128: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
129: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
130: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
131: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
132: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
133: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
134: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
135: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
136: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
137: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
138: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
139: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
140: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
141: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
142: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
143: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
144: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
145: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
146: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
147: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
148: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
149: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
150: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
151: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
152: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
153: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
154: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
155: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
156: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
157: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
158: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
159: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
160: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
161: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
162: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
163: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
164: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
165: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
166: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
167: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
168: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
169: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
170: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
171: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
172: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
173: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
174: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
(K){20}
175: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
176: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
177: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
178: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
179: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
180: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
181: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
182: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
183: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
184: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
185: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
186: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
187: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
188: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
189: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
190: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
191: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
192: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
193: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
194: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
195: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
196: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
197: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
198: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
199: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
200: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
201: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
202: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
203: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
204: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
205: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
206: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
207: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
208: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
209: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
210: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
211: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
212: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
213: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
214: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
215: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
216: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
217: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
218: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
219: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
220: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
221: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
222: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
223: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKN
                (K){20}
```

```
  224: KKKKK KKKKKKKKKKKKKKKKKKKK KKKNR

              (K){20}
  225: KKKKK KKKKKKKKKKKKKKKKKKKK KKNRH

              (K){20}
  226: KKKKK KKKKKKKKKKKKKKKKKKKK KNRHH

              (K){20}
  227: KKKKK KKKKKKKKKKKKKKKKKKKK NRHHI


     Q64075  ck: 8048  len: 215   ! Q64075 rattus sp. nucleoporin p62 homolog p
1            (R,K){20,20}
              (K){20}
   35: CEFLE KKKKKKKKKKKKKKKKKKKK KKTGD

              (K){20}
   36: EFLEK KKKKKKKKKKKKKKKKKKKK KTGDN

              (K){20}
   37: FLEKK KKKKKKKKKKKKKKKKKKKK TGDNA


     Q9D5G1  ck: 9388  len: 169   ! Q9d5g1 mus musculus (mouse). adult male tes
1            (R,K){20,20}
              (R){20}
  117: VQLRG RRRRRRRRRRRRRRRRRRRR RRKKK

              (R){20}
  118: QLRGR RRRRRRRRRRRRRRRRRRRR RKKKE

              (R){20}
  119: LRGRR RRRRRRRRRRRRRRRRRRRR KKKEE

             (R,K){20}
  120: RGRRR RRRRRRRRRRRRRRRRRRRK KKEEE

             (R,K){20}
  121: GRRRR RRRRRRRRRRRRRRRRRRKK KEEEE

             (R,K){20}
  122: RRRRR RRRRRRRRRRRRRRRRRKKK EEEEE


     O35807  ck: 7510  len: 129   ! O35807 rattus norvegicus (rat). microvascul
1            (R,K){20,20}
              (K){20}
   85: VLLAS KKKKKKKKKKKKKKKKKKKK KIKWE

              (K){20}
   86: LLASK KKKKKKKKKKKKKKKKKKKK IKWEG


     Q8BXG9  ck: 5434  len: 115   ! Q8bxg9 mus musculus (mouse). hypothetical a
1            (R,K){20,20}
             (R,K){20}
   40: IIIII RRRRRRRRRRRRRRRRRRRK KETGS

             (R,K){20}
   41: IIIIR RRRRRRRRRRRRRRRRRRKK ETGSH


     Q8BHV2  ck: 8958  len: 154   ! Q8bhv2 mus musculus (mouse). weakly similar
1            (R,K){20,20}
              (R){20}
   42: RRGEE RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
   43: RGEER RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
   44: GEERR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
   45: EERRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
   46: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
   47: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK

              (R){20}
   48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKR

              (R){20}
   49: RRRRR RRRRRRRRRRRRRRRRRRRR RRKRE

              (R){20}
   50: RRRRR RRRRRRRRRRRRRRRRRRRR RKRER

              (R){20}
   51: RRRRR RRRRRRRRRRRRRRRRRRRR KRERE

             (R,K){20}
   52: RRRRR RRRRRRRRRRRRRRRRRRRK RERER

             (R,K){20}
   53: RRRRR RRRRRRRRRRRRRRRRRRKR ERERE

Databases searched:
      SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
      SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

   Total finds:           348
  Total length:   305,079,309
Total sequences:      958,388
      CPU time:      11:25.82
```

```
!!SEQUENCE_LIST 1.0
! FINDPATTERNS on swp:* allowing 0 mismatches

!       1 (R,K){20,20}                          January 30, 2004 07:59 ..

SP_FUN:Q12444       ck: 1384  len: 126   finds: 3    ! Q12444 saccharomyces cerevisia
SP_FUN:Q9P529       ck: 291   len: 128   finds: 13   ! Q9p529 neurospora crassa. hypo
SP_HUM:Q9NT34       ck: 7330  len: 380   finds: 7    ! Q9nt34 homo sapiens (human). h
SP_HUM:Q9H6Q7       ck: 3351  len: 720   finds: 10   ! Q9h6q7 homo sapiens (human). h
SP_HUM:Q8N6F0       ck: 9898  len: 55    finds: 15   ! Q8n6f0 homo sapiens (human). s
SP_HUM:Q9HC48       ck: 7602  len: 667   finds: 1    ! Q9hc48 homo sapiens (human). c
SP_HUM:Q9H5V6       ck: 379   len: 168   finds: 10   ! Q9h5v6 homo sapiens (human). h
SP_IN:Q8SWR7        ck: 2372  len: 515   finds: 4    ! Q8swr7 drosophila melanogaster
SP_IN:Q8T2U7        ck: 8768  len: 791   finds: 2    ! Q8t2u7 dictyostelium discoideu
SP_IN:Q8I247        ck: 5951  len: 206   finds: 2    ! Q8i247 plasmodium falciparum (
SP_OM:Q95LV6        ck: 7515  len: 531   finds: 11   ! Q95lv6 macaca fascicularis (cr
SP_PL:Q9LXR2        ck: 4143  len: 517   finds: 4    ! Q9lxr2 arabidopsis thaliana (m
SP_PL:Q8S7D3        ck: 6479  len: 80    finds: 10   ! Q8s7d3 oryza sativa (rice). hy
SP_PL:Q8LQP6        ck: 3239  len: 113   finds: 9    ! Q8lqp6 oryza sativa (japonica
SP_PL:Q9LGZ9        ck: 6094  len: 260   finds: 222  ! Q9lgz9 arabidopsis thaliana (m
SP_RO:Q64075        ck: 8048  len: 215   finds: 3    ! Q64075 rattus sp. nucleoporin
SP_RO:Q9D5G1        ck: 9388  len: 169   finds: 6    ! Q9d5g1 mus musculus (mouse). a
SP_RO:O35807        ck: 7510  len: 129   finds: 2    ! O35807 rattus norvegicus (rat)
SP_RO:Q8BXG9        ck: 5434  len: 115   finds: 2    ! Q8bxg9 mus musculus (mouse). h
SP_RO:Q8BHV2        ck: 8958  len: 154   finds: 12   ! Q8bhv2 mus musculus (mouse). w
\\End of list

 Databases searched:
        SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
        SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

     Total finds:            348
    Total length:    305,079,309
 Total sequences:        958,388
        CPU time:       12:54.92
```

```
!!AA_SEQUENCE 1.0
ID   Q12444       PRELIMINARY;      PRT;   126 AA.
AC   Q12444;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   ORF YOR309C.
GN   YOR309C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pearson B.M., Hernando Y., Kalogeropoulos A., Schweizer M.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   MIPS;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FY1679;
RA   Pearson B.M., Hernando Y., Wolf S.S., Kalogeropoulos A., Schweizer M.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z75217; CAA99629.1; -.
DR   EMBL; X90565; CAA62164.1; -.
DR   SGD; S0005836; YOR309C.
SQ   SEQUENCE   126 AA;  16294 MW;  46EF1F4C664802C8 CRC64;

 Q12444  Length: 126  January 30, 2004 08:18  Type: P  Check: 1384  ..

      1  MQMLIPQRLL LILNPILMMK RKKRKKRKKR RERETMMKIP RILKKLRRKR

     51  RTRRKRRKRR KRRRRKRRKR RRKRSPRKRR KRRNKDAFYI LIISDPSRSL

    101  LFGFRKFSII IQCLTYFSFH ILFHNL

!!AA_SEQUENCE 1.0
ID   Q9P529       PRELIMINARY;      PRT;   128 AA.
AC   Q9P529;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update) .
DE   Hypothetical 15.2 kDa protein.
GN   B24H17.160.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356815; CAB92638.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   128 AA;  15157 MW;  8C7C65C3DFB70765 CRC64;

  Q9P529  Length: 128  January 30, 2004 08:18  Type: P  Check: 291  ..

      1  MAIISIGLHH KNINRRAPGH SVYSKNSSYD FQQYDATQHY LPSQGFKAIP

     51  DHLTGGKDGC LSTHDKRKNQ KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK

    101  KKEQESRTYF QQHFQADGIC PTPEWHTR

!!AA_SEQUENCE 1.0
ID   Q9NT34       PRELIMINARY;      PRT;   380 AA.
AC   Q9NT34;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP434I1120.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137556; CAB70810.1; -.
DR   Genew; HGNC:15736; C17orf28.
KW   Hypothetical protein.
FT   NON_TER     380    380
SQ   SEQUENCE   380 AA;  42689 MW;  67F50DD101346AFB CRC64;

 Q9NT34  Length: 380  January 30, 2004 08:18  Type: P  Check: 7330  ..

      1  MGSTDSKLNF RKAVIQLTTK TQPVEATDDA FWDQFWADTA TSVQDVFALV

     51  PAAEIRAVRE ESPSNLATLC YKAVEKLVQG AESGCHSEKE KQIVLNCSRL

    101  LTRVLPYIFE DPDWRGFFWS TVPGAGRGGG EEDDEHARPL AESLLLAIAD

    151  LLFCPDFTVQ SHRRSTVDSA EDVHSLDSCE YIWEAGVGFA HSPQPNYIHD

    201  MNRMELLKLL LTCFSEAMYL PPAPESGSTN PWVQFFCSTE NRHALPLFTS

    251  LLNTVCAYDP VGYGIPYNHL LFSDYREPLV EEAAQVLIVT LDHDSASSAS

    301  PTVDGTTTGT AMDDADPPGP ENLFVNYLSR IHREEDFQFI LKGIARLLSN

    351  LLLQKKKKKK KKKKKKKKKK KKKKKKKKKK

!!AA_SEQUENCE 1.0
ID   Q9H6Q7       PRELIMINARY;      PRT;   720 AA.
AC   Q9H6Q7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ21979 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK025632; BAB15196.1; -.
KW   Hypothetical protein.
FT   NON_TER     720    720
SQ   SEQUENCE   720 AA;  84029 MW;  A86586FEAA953D0B CRC64;

 Q9H6Q7  Length: 720  January 30, 2004 08:18  Type: P  Check: 3351  ..

      1  MLTEQVEQYT KEMEKNTCII EDLKNELQRN KGASTLSQQT HMKIQSTLDI

     51  LKEKTKEAER TAELAEADAR EKDKELVEAL KRLKDYESGV YGLEDAVVEI

    101  KNCKNQIKIR DREIEILTKE INKLELKISD FLDENEALRE RVGLEPKTMI

    151  DLTEFRNSKH LKQQQYRAEN QILLKEIESL EEERLDLKKK IRQMAQERGK

    201  RSATSGLTTE DLNLTENISQ GDRISERKLD LLSLKNMSEA QSKNEFLSRE
```

```
    251  LIEKERDLER SRTVIAKFQN KLKELVEENK QLEEGMKEIL QAIKEMQKDP

    301  DVKGGETSLI IPSLERLVNA IESKNAEGIF DASLHLKAQV DQLTGRNEEL

    351  RQELRESRKE AINYSQQLAK ANLKIDHLEK ETSLLRQSEG SNVVFKGIDL

    401  PDGIAPSSAS IINSQNEYLI HLLQELENKE KKLKNLEDSL EDYNRKFAVI

    451  RHQQSLLYKE YLSEKETWKT ESKTIKEEKR KLEDQVQQDA IKVKEYNNLL

    501  NALQMDSDEM KKILAENSRK ITVLQVNEKS LIRQYTTLVE LERQLRKENE

    551  KQKNELLSME AEVCEKIGCL QRFKEMAIFK IAALQKVVDN SVSLSELELA

    601  NKQYNELTAK YRDILQKDNM LVQRTSNLEH LECENISLKE QVESINKELE

    651  ITKEKLHTIE QAWEQETKLG NESSMDKAKK SITNSDIVSI SKKKKKKKKK

    701  KKKKKKKKKK KKKKKKKKKK

!!AA_SEQUENCE 1.0
ID   Q8N6F0      PRELIMINARY;      PRT;    55 AA.
AC   Q8N6F0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to LOC201361.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC030525; AAH30525.1; -.
SQ   SEQUENCE   55 AA;  7251 MW;  0906032B284006BA CRC64;

  Q8N6F0  Length: 55  January 30, 2004 08:18  Type: P  Check: 9898  ..

       1  MRLASQRRRR GRGKKKKRKR KKKKKKKKKK KKKKKKKKKK KRRRRRRGRR

      51  RRMQQ

!!AA_SEQUENCE 1.0
ID   Q9HC48      PRELIMINARY;      PRT;   667 AA.
AC   Q9HC48;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CTCL tumor antigen se2-5 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=21143360; PubMed=11149944;
RA   Eichmuller S., Usener D., Dummer R., Stein A., Thiel D.,
RA   Schadendorf D.;
RT   "Serological detection of cutaneous T-cell lymphoma-associated
RT   antigens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:629-634(2001).
CC   -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
DR   EMBL; AF177228; AAG33676.1; -.
DR   HSSP; Q12923; 3PDZ.
DR   InterPro; IPR001478; PDZ.
DR   Pfam; PF00595; PDZ; 2.
DR   SMART; SM00228; PDZ; 2.
DR   PROSITE; PS50106; PDZ; 2.
FT   NON_TER       1      1
FT   NON_TER     667    667
SQ   SEQUENCE   667 AA;  73499 MW;  C653EC16802BAE02 CRC64;

  Q9HC48  Length: 667  January 30, 2004 08:18  Type: P  Check: 7602  ..

       1  EHENLFREND CIVRINDGDL RNRRFEQAQH MFRQAMRTPI IWFHVVPAAN

      51  KEQYEQLSQS EKNNYYSSRF SPDSQYIDNR SVNSAGLHTV QRAPRLNHPP

     101  EQIDSHSRLP HSAHPSGKPP SAPASAPQNV FSTTVSSGYN TKKIGKRLNI

     151  QLKKGTEGLE FSITSRDVTI GGSAPIYVKN ILPRGAAIQD GRLKAGDRLI

     201  EVNGVDLVGK SQEEVVSLLR STKMEGTVSL LVFRQEDAFH PRELNAEPSQ

     251  MQIPKETKAE DEDIVLTPDG TREFLTFEVP LNDSGSAGLG VSVKGNRSKE

     301  NHADLGIFVK SIINGGAASK DGRLRVNDQL IAVNGESLLG KTNQDAMETL

     351  RRSMSTEGNK RGMIQLIVAR RISKCNELKS PGSPPGPELP IETALDDRER

     401  RISHSLYSGI EGLDESPSRN AALSRIMGES GKYQLSPTVN MPQDDTVIIE

     451  DDRLPVLPPH LSDQSSSSSH DDVGFVTADA GTWAKAAISD SADCSLSPDV

     501  DPVLAFQREG FGRQIADETK LNTVDDQKAG SPSRDVGPSL GLKKSSSLES

     551  LQTAVAEVTL NGDIPFHRPR PRIIRGRGCN ESFRAAIDKS YDKPAVDDDD

     601  EGMETLEEDT EESSRSGRES VSTASDQPSH SLERQMNGNQ EKGDKTDRKK

     651  KKKKKKKKKK KKKKKKK

!!AA_SEQUENCE 1.0
ID   Q9H5V6      PRELIMINARY;      PRT;   168 AA.
AC   Q9H5V6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ22976 (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA   Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
RA   Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK026629; BAB15513.1; -.
KW   Hypothetical protein.
FT   NON_TER     168    168
SQ   SEQUENCE   168 AA;  19549 MW;  A19DBD195F8A1A90 CRC64;

  Q9H5V6  Length: 168  January 30, 2004 08:18  Type: P  Check: 379  ..

       1  MGNGGRSGLQ QGKGNVDGVA ATPTAASASC QYRCIECNQE AKELYRDYNH

      51  GVLKITICKS CQKPVDKYIE YDPVIILINA ILCKAQAYRH ILFNTQINIH

     101  GKLCIFCLLC EAYLRWWQLQ DSNQNTAPDD LIRYVREWEK KKKKKKKKKK

     151  KKKKKKKKKK KKKKKKK

!!AA_SEQUENCE 1.0
ID   Q8SWR7      PRELIMINARY;      PRT;   515 AA.
AC   Q8SWR7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
```

```
DE   GH22607p (Fragment).
GN   CG7180.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY095518; AAM12251.1; -.
DR   FlyBase; FBgn0032673; CG7180.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_PP.
DR   Pfam; PF00102; Y_phosphatase; 1.
DR   SMART; SM00194; PTPc; 1.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW   Hydrolase.
FT   NON_TER     515    515
SQ   SEQUENCE   515 AA;  59080 MW;  B2825B7EEA96195E CRC64;

 Q8SWR7  Length: 515  January 30, 2004 08:18  Type: P  Check: 2372  ..

       1  MMTTQQLVGK PEAAENEEKS PAAASAAAGA GADMTAATGG GSSGAGGGKT

      51  GKHRSRSSAR YDDVEKQQRK SRAIVSIPNT IKLSMLNSGL LSFERIKLEA

     101  RDENNSLSKT IPNGPIDIRH FLKLCDLRRK FPVLYKLEFQ TAAKVESNTC

     151  RHALKKNNLE KNQNPKCIPY DYNRVVLEKV GGLQDSDYVN ASYVDSLLKP

     201  NAYIVTQGPV EETVQAYWRM VWQENISAIV MLTKTFDFAK VMCHQYWPPN

     251  MEVHEQYGDI FINIVREEQL ANFHIRTFRL YKMNEKQEVT DERLILQFHY

     301  TEWYSHSCPF SNALLEFRRR VRLVVGNIIK DEDDMRGPIL VHCSDGGGRS

     351  GVYMSIDANL ELAEEEECFN VFGYLKKLRQ SRKGLVENVE QYKFIYDTLE

     401  EHIICGKTWF PVSELSDRLK AKARRNSGTK MNEYQAEYDQ ICKQTPRFTI

     451  GDCAGGHRAD NREKNRDVLC VPPDNFRPYL TSFQGNAFTD YIKKKKKKKK

     501  KKKKKKKKKK KKKKK

!!AA_SEQUENCE 1.0
ID   Q8T2U7      PRELIMINARY;      PRT;   791 AA.
AC   Q8T2U7;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 92.4 kDa protein.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115574; AAL92183.1; -.
DR   InterPro; IPR005033; YEATS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF03366; YEATS; 1.
DR   SMART; SM00355; ZnF_C2H2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   791 AA;  92375 MW;  D66CCB6DEC92352C CRC64;

 Q8T2U7  Length: 791  January 30, 2004 08:18  Type: P  Check: 8768  ..

       1  MKEKIETKLL DTIDEKIENS TTTTTKNNTH NNNNTNTNNN NNNNNNNNNN

      51  NNNNNNNNNN NNNNNNNNNN NNPFSTLSAS LSKLKRKLII NNEGDREELK

     101  KRLITMIERE FPAELKNKEQ ELIEIDDRIN TVKEMLNSLN KQRNSTKSST

     151  YHIPHTNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

     201  NNKPKRIFHK LQNGEFVKLV CKYCRRSDFV SGLGFLNHVR IKHGFRYSTL

     251  DEAAEQIGVP VADSEIPKDD PSREGIVFRF KRGAYSPSIA FERNSEEDNG

     301  TFDNNNNNNN NNNNNNNNNN NNNNNNNTGN DTDKKKNNNG NDADVDMIDL

     351  NVNSNYSNNK ETNEESSGSS RFYVKKKIIV GNTSTQIHPD YRGHDRSTHK

     401  WTVYVRGPQN EADISYFVKK IWFYLHDSFA PNDKVEVVER PFNLTRRGWG

     451  EFPVRIRLFF HDKRNKPIDI IHNLKLIQLP IQYNVPVVGG ETTTEIDLDR

     501  LFFDKREKQQ KLDQLNNNNN NNNNSENNNN NNNNNNNNNN NNINNNNNNN

     551  INNNNNDNNN SSNNTSPTSN YLNEPQKIVN NDKVSENSSN NEDSQKNKEK

     601  EKEKEKEKEK EKEKEKEKEK EKEKEKDKDK EKEKDKEKDK VKERDNKEKE

     651  LENNRPKDDR DRDREKERKE KGKGKGKDIE IEIEIETDIG IETEIGIETE

     701  TETEIETGKE KEKEKEKEKE KEKEKEKEKG KEIEIGRERG KGKGKEKEIE

     751  TEMEIGKEIE AEIEIEKEKK KKKKKKKKKK KKKKKKKKKE I

!!AA_SEQUENCE 1.0
ID   Q8I247      PRELIMINARY;      PRT;   206 AA.
AC   Q8I247;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   PFA0475C.
OS   Plasmodium falciparum (isolate 3D7).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=36329;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22255708; PubMed=12368867;
RA   Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA   Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA   Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA   Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA   Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA   Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA   Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA   Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA   Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA   Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA   Oliver K., Ormond D., Price C., Quail M.A., Rabbinowitsch E.,
RA   Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA   Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA   Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA   Sulston J.E., Craig A., Newbold C., Barrell B.G;
RT   "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL   Nature 419:527-531(2002).
DR   EMBL; AL031745; CAD49055.1; -.
```

```
KW   Hypothetical protein.
SQ   SEQUENCE   206 AA;  25047 MW;  1192E49A3DC4523F CRC64;

 Q8I247  Length: 206  January 30, 2004 08:18  Type: P  Check: 5951  ..

       1  MEGQHKEKNT KIKKSKKPLV VSNRKPFNVF EKKKSAKPLV RDPRFSDFSG

      51  SFNANFFRNA YKFLYDSREQ EKKIIEKKLK SKNITQEEKD ELKKKYNDYK

     101  STDILLKKKE EERKLKAELV KQEKQNILTK NKKPYYYSDR KIKKIVQEKL

     151  SSYKSLKKVI KKEKKTLQKE RKRNIKPTKK KIFLKKKKKK KKKKKKKKKK

     201  KKKKKT

!!AA_SEQUENCE 1.0
ID   Q95LV6      PRELIMINARY;      PRT;   531 AA.
AC   Q95LV6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 61.4 kDa protein (Fragment).
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA   Terao K., Sugano S.;
RT   "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT   libraries.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB071085; BAB64479.1; -.
KW   Hypothetical protein.
FT   NON_TER     531    531
SQ   SEQUENCE   531 AA;  61389 MW;  B55996B4F5CDD60C CRC64;

 Q95LV6  Length: 531  January 30, 2004 08:18  Type: P  Check: 7515  ..

       1  MESESSNAMN MNVQHEREDK.NIQKMLPESV PCYSQHLSFS TYQMKDPDPC

      51  KSRSEPKSPE GRSSWNLSHI VQKTEQETHF RESVLEPISG YMMKQSPHMQ

     101  EGIKCVVGLK TSFPKTGKSE IGSMPHDTPW DENPRRKWDS SISEKTAWNP

     151  KNLQTVLKPL DFSSLMSSEY ESRSYTLEFI GKKSITSPKH VTLKTKQLPI

     201  SQLFNIIRCS TENHRKKKQH CFKYKMKGRQ WYTSIGEALR SATEYAKSPP

     251  SKSMIDKLLF NTAARGTLSN RTHHQNVYGH TTEEKEEVQE NVAASSLGPL

     301  DFFMPVLSDS KNQTNTIRLS ERKTILNPKC LTMKEKKSPI SQIRKINRHF

     351  TTKHKKKLES NLKTKLKAMW QGENVTDTFP NTISFTPDTS DIKRQSRFQT

     401  EIDMRISGLS HTQPTQIESL AEGIARCSDK RRTSNLVKGT KLHDRESGEK

     451  KQEHLTGMDP FYAENFMTNT HLRKDPHLGK SEDVLLGEPF ISKSQFYKGN

     501  SKKKKKKKKK KKKKKKKKKK KKKKKKKKKK K

!!AA_SEQUENCE 1.0
ID   Q9LXR2      PRELIMINARY;      PRT;   517 AA.
AC   Q9LXR2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 59.7 kDa protein.
GN   T20N10_250.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,
RA   Rudd S., Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353032; CAB88307.1; -.
DR   InterPro; IPR001810; F-box.
DR   InterPro; IPR006566; FBD.
DR   Pfam; PF00646; F-box; 1.
DR   SMART; SM00579; FBD; 1.
DR   SMART; SM00256; FBOX; 1.
DR   PROSITE; PS50181; FBOX; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   517 AA;  59689 MW;  EC6D957D01F86E70 CRC64;

 Q9LXR2  Length: 517  January 30, 2004 08:18  Type: P  Check: 4143  ..

       1  MDLFSSLPNE LLYHILSFLS TKEAALTSVL SKRWRNLFAF VPYLEFDDSV

      51  FLHPEERKRE KEGILQSFMD FVDRVLDLHG DSLIKTFSLK CKTGVDSDHV

     101  DRWICNVLAR GVSDLDLFID FRDLYSLPHE VGVSRTLVVL RVGSESDLYW

     151  WQKFLCLPML KTLVLDSCWL CIGQFQILLL ACPALEELDM TNTRWKDSNV

     201  TVSSSILKEL TIDLHGCCSV VNLKSLSFDA PSLVYFYYCD SLAEDYPQVN

     251  LKNLVEAQIN LLLTQAQIEQ VRALNNEMLV ADDVFPGLGN AWKLITGLRN

     301  VQQLYLSPDT LEVLSRCCEG MPVFNNLKVL SIWSDMNRGW QAMPVLLRNC

     351  PHLETLIIEG LLHYATDKCG DVCDCISRDY KDHSLTSCPV KKLQIYEFRG

     401  TIRELEMIKH FLKIFPCLKE MDIYAHENSH TLFKDPTIFE RVGKKKKKKK

     451  KKKKKKKKKK KKKKKKIRLN FKPVNKTEQF LKRLADKLCF IPQCLEFLDV

     501  DSSLGELALL AMDSRPS

!!AA_SEQUENCE 1.0
ID   Q8S7D3      PRELIMINARY;      PRT;    80 AA.
AC   Q8S7D3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical 9.4 kDa protein.
GN   OSJNBA0057L21.23.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA   Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA   Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pai G.,
RA   VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA   Salzberg S.L., White O., Fraser C.M.;
RT   "Oryza sativa chromosome 10 BAC OSJNBa0057L21 genomic sequence.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC087599; AAL79706.1; -.
DR   Gramene; Q8S7D3; -.
KW   Hypothetical protein.
```

```
SQ   SEQUENCE   80 AA;  9362 MW;  0177C863133B21D8 CRC64;

  Q8S7D3  Length: 80  January 30, 2004 08:18  Type: P  Check: 6479  ..

       1  MAVTMRQPVG RRWSEAAGGA TPGGRWGHWW WPATAATRVF LSVIHLDKKK

      51  KKKKKKKKKK KKKKKKKKKK KKKKKKLLGE

!!AA_SEQUENCE 1.0
ID   Q8LQP6      PRELIMINARY;      PRT;   113 AA.
AC   Q8LQP6;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   OJ1117_G01.13 protein.
GN   OJ1117_G01.13.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:OJ1117_G01.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003374; BAB93330.1; -.
DR   Gramene; Q8LQP6; -.
SQ   SEQUENCE   113 AA;  13660 MW;  597DB0EDEB2AA3EF CRC64;

 Q8LQP6  Length: 113  January 30, 2004 08:18  Type: P  Check: 3239  ..

       1  MEATSLEHIK KKKKKKKKKK KKKKKKKKKK KKKKKKRERE REEEDEEEEE

      51  EEEELKLNEY IWDIYEAAKG INEKLMPGIV VYLAVHDECV RLRLTGVDVK

     101  HEAWEICLTP VQV

!!AA_SEQUENCE 1.0
ID   Q9LGZ9      PRELIMINARY;      PRT;   260 AA.
AC   Q9LGZ9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Genomic DNA, chromosome 3, BAC clone:F1D9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RA   Nakamura Y.;
RT   "Structural Analysis of Arabidopsis thaliana Chromosome 3. III.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002460; BAA97098.1; -.
DR   InterPro; IPR005819; Histone_H5.
DR   PRINTS; PR00624; HISTONEH5.
SQ   SEQUENCE   260 AA;  33307 MW;  43E2394CB8131143 CRC64;

 Q9LGZ9  Length: 260  January 30, 2004 08:18  Type: P  Check: 6094  ..

       1  MDRCIRKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK

      51  KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK

     101  KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK

     151  KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK

     201  KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKKKKK KKKKKKNRHH

     251  INWVMGFVIF

!!AA_SEQUENCE 1.0
ID   Q64075      PRELIMINARY;      PRT;   215 AA.
AC   Q64075;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Nucleoporin p62 homolog protein (Fragment).
OS   Rattus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10118;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95151924; PubMed=7849178;
RA   Wang Z.Q., Akmal K.M., Kim K.H.;
RT   "An unusual nucleoporin-related messenger ribonucleic acid is present
RT   in the germ cells of rat testis.";
RL   Biol. Reprod. 51:1022-1030(1994).
DR   EMBL; S75997; AAB33384.1; -.
KW   Porin.
FT   NON_TER       1      1
SQ   SEQUENCE   215 AA;  24593 MW;  098251C97A8FBD88 CRC64;

 Q64075  Length: 215  January 30, 2004 08:18  Type: P  Check: 8048  ..

       1  SGGRATSSCD EDCLSSSLPF SLSGPVKQDC EFLEKKKKKK KKKKKKKKKK

      51  KKKKKKTGDN AKSVSRQYSL KVTKLEHEAE QAKVELDFIL SQQKELEDLL

     101  SPLEESVKEQ SGTIYLQHAD EEREKTYKLA ENIDAQLKRM AQDLKDIIEH

     151  LNMAGGPADT SDPLQQICKI LNAHMDSLQW VDQSSALLQR RVEEASRVCE

     201  SRRKEQERSL RIAFD

!!AA_SEQUENCE 1.0
ID   Q9D5G1      PRELIMINARY;      PRT;   169 AA.
AC   Q9D5G1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Adult male testis cDNA, RIKEN full-length enriched library,
DE   clone:4930444P10 product:hypothetical Arginine-rich region containing
DE   protein, full insert sequence (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
```

```
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
DR   EMBL; AK015384; BAB29822.2; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   169 AA;  19305 MW;  91B9959380A694CC CRC64;

 Q9D5G1  Length: 169  January 30, 2004 08:18  Type: P  Check: 9388  ..

       1  RCTGQAGPQL RALAGPGWPR LAPALLSGGR ARNIAGLPAA KHARPESGAS

      51  ARRLPAAPPH RGGQPGDAAS LSRELASTHG RRWLPHCPLP PLAQTSSLVP

     101  WVRHLKKPIS SVQLRGRRRR RRRRRRRRRR RRRRRRRRKK KEEEEEEEVA

     151  FLDSLIEPRL TSNLLHRHE

!!AA_SEQUENCE 1.0
ID   O35807      PRELIMINARY;      PRT;   129 AA.
AC   O35807;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   MICROVASCULAR endothelial differentiation protein 2.
GN   MDG2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Epididymis;
RX   MEDLINE=98172708; PubMed=9511718;
RA   Proels F., Loser B., Marx M.;
RT   "Differential expression of osteopontin, PC4, and CEC5, a novel mRNA
RT   species, during in vitro angiogenesis.";
RL   Exp. Cell Res. 239:1-10(1998).
DR   EMBL; Y08769; CAA70022.1; -.
DR   InterPro; IPR000719; Prot_kinase.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   ATP-binding; Transferase.
SQ   SEQUENCE   129 AA;  15080 MW;  38102272BBE2EDB4 CRC64;

 O35807  Length: 129  January 30, 2004 08:18  Type: P  Check: 7510  ..

       1  MLKHPHIVEL LETYSSDGML YMVFEFMDGA DLCFEIVKRA DAGFVYSEAV

      51  ASHYMRQILE ALRYCHDNNI IHRDVKPHCV LLASKKKKKK KKKKKKKKKK

     101  KKKKKIKWEG RDAFWEAIPV KSSGGVIIQ

!!AA_SEQUENCE 1.0
ID   Q8BXG9      PRELIMINARY;      PRT;   115 AA.
AC   Q8BXG9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical arginine-rich region containing protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK047167; BAC32979.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   115 AA;  13938 MW;  6C0F7EED8669CF65 CRC64;

 Q8BXG9  Length: 115  January 30, 2004 08:18  Type: P  Check: 5434  ..

       1  GRGRRISEFE VSQGYTEKPC LEKPKKKKKK IIIIIIIIR RRRRRRRRRR

      51  RRRRRRRRKK ETGSHFVALA SLELHLPPEC WNLRYRLPHP TQIGFNAIKI

     101  SIKDTSLSLL AFCVK

!!AA_SEQUENCE 1.0
ID   Q8BHV2      PRELIMINARY;      PRT;   154 AA.
AC   Q8BHV2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Weakly similar to hypothetical 10.3 kDa protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK082253; BAC38447.1; -.
KW   Hypothetical protein.
```

```
SQ   SEQUENCE   154 AA;  19596 MW;  AEE8158A4839A216 CRC64;
 Q8BHV2  Length: 154  January 30, 2004 08:18  Type: P  Check: 8958  ..
       1  MVQESRKQSN VGQKTSRRRN VFKERVRFSR VPRDRERRGE ERRRRRRRRR
      51  RRRRRRRRRR RRRRRRRRRR KRERERESEH EVSNANKDFG LLWIHFFCAC
     101  DFMGRIFISH QRKRNMAFLP SGDIIDRYLS YNWVMRSLPS ILAYYMLKHC
     151  GGCT
```

```
! FINDPATTERNS on geneseqp:* allowing 0 mismatches
!       1 (R,K){20,20}                          January 30, 2004 07:18 ..

         AAP20159  ck: 5750  len: 20    ! Aap20159 Sequence of lysine polymer. 8/2002

1                  (R,K){20,20}
                     (K){20}
          1:       KKKKKKKKKKKKKKKKKKKK

         AAP61030  ck: 9157  len: 898   ! Aap61030 Entire coded sequence from plasmid

1                  (R,K){20,20}
                     (K){20}
        873: KNITW KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        874: NITWK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        875: ITWKK KKKKKKKKKKKKKKKKKKKK KKKK

                     (K){20}
        876: TWKKK KKKKKKKKKKKKKKKKKKKK KKK

                     (K){20}
        877: WKKKK KKKKKKKKKKKKKKKKKKKK KK

                     (K){20}
        878: KKKKK KKKKKKKKKKKKKKKKKKKK K

                     (K){20}
        879: KKKKK KKKKKKKKKKKKKKKKKKKK

         AAP61056  ck: 2017  len: 899   ! Aap61056 Translation of plasmid pAU157 encc

1                  (R,K){20,20}
                     (K){20}
        873: KNITW KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        874: NITWK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        875: ITWKK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        876: TWKKK KKKKKKKKKKKKKKKKKKKK KKKK

                     (K){20}
        877: WKKKK KKKKKKKKKKKKKKKKKKKK KKK

                     (K){20}
        878: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                     (K){20}
        879: KKKKK KKKKKKKKKKKKKKKKKKKK K

                     (K){20}
        880: KKKKK KKKKKKKKKKKKKKKKKKKK

         AAP61082  ck: 7915  len: 898   ! Aap61082 Complete translation of plasmid pA

1                  (R,K){20,20}
                     (K){20}
        873: KNITW KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        874: NITWK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        875: ITWKK KKKKKKKKKKKKKKKKKKKK KKKK

                     (K){20}
        876: TWKKK KKKKKKKKKKKKKKKKKKKK KKK

                     (K){20}
        877: WKKKK KKKKKKKKKKKKKKKKKKKK KK

                     (K){20}
        878: KKKKK KKKKKKKKKKKKKKKKKKKK K

                     (K){20}
        879: KKKKK KKKKKKKKKKKKKKKKKKKK

         AAR29580  ck: 4341  len: 657   ! Aar29580 FMR-1 gene product. 3/2003

1                  (R,K){20,20}
                     (R){20}
         19: RRRRP RRRRRRRRRRRRRRRRRRRR RRRRL

                     (R){20}
         20: RRRPR RRRRRRRRRRRRRRRRRRRR RRRLG

                     (R){20}
         21: RRPRR RRRRRRRRRRRRRRRRRRRR RRLGL

                     (R){20}
         22: RPRRR RRRRRRRRRRRRRRRRRRRR RLGLE

                     (R){20}
         23: PRRRR RRRRRRRRRRRRRRRRRRRR LGLER

         AAW03642  ck: 9623  len: 116   ! Aaw03642 Human cannabinoid GPR N-terminal s

1                  (R,K){20,20}
                     (K){20}
         34: QYEDI KKKKKKKKKKKKKKKKKKKK KSPFQ

                     (K){20}
         35: YEDIK KKKKKKKKKKKKKKKKKKKK SPFQE

         AAW38839  ck: 801   len: 28    ! Aaw38839 Delivery peptide used in peptide m

1                  (R,K){20,20}
                     (K){20}
          1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
          2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
          3:    KK KKKKKKKKKKKKKKKKKKKK KKKKX

                     (K){20}
          4:   KKK KKKKKKKKKKKKKKKKKKKK KKKXK

                     (K){20}
          5:  KKKK KKKKKKKKKKKKKKKKKKKK KKXK

                     (K){20}
          6: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

                     (K){20}
          7: KKKKK KKKKKKKKKKKKKKKKKKKK XK

         AAW38840  ck: 2989  len: 29    ! Aaw38840 Delivery peptide used in peptide m
```

```
1              (R,K){20,20}
                 (K){20}
      1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
      6: KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
      7: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
      8: KKKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38841  ck: 5252  len: 30    ! Aaw38841 Delivery peptide used in peptide m
1              (R,K){20,20}
                 (K){20}
      1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
      6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
      7: KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
      8: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
      9: KKKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38842  ck: 7590  len: 31    ! Aaw38842 Delivery peptide used in peptide m
1              (R,K){20,20}
                 (K){20}
      1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}

      5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
      7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
      8: KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
      9: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
     10: KKKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38877  ck: 1129  len: 23    ! Aaw38877 Delivery peptide used in peptide m
1              (R,K){20,20}
                 (K){20}
      1:       KKKKKKKKKKKKKKKKKKKK VTK

AAW38843  ck: 3     len: 32    ! Aaw38843 Delivery peptide used in peptide m.
1              (R,K){20,20}
                 (K){20}
      1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
      8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
      9: KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
     10: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
     11: KKKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38878  ck: 2949  len: 24    ! Aaw38878 Delivery peptide used in peptide m
1              (R,K){20,20}
                 (K){20}
      1:       KKKKKKKKKKKKKKKKKKKK KVTK

                 (K){20}
      2:     K KKKKKKKKKKKKKKKKKKKK VTK

AAW38844  ck: 2491  len: 33    ! Aaw38844 Delivery peptide used in peptide ma
```

```
1            (R,K){20,20}
               (K){20}
     1:        KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     2:      K KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     3:     KK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     4:    KKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     5:   KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     6:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     7:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     8:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

               (K){20}
     9:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

               (K){20}
    10:  KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

               (K){20}
    11:  KKKKK KKKKKKKKKKKKKKKKKKKK KXK

               (K){20}
    12:  KKKKK KKKKKKKKKKKKKKKKKKKK XK

    AAW38879  ck: 4844  len: 25   ! Aaw38879 Delivery peptide used in peptide m

1            (R,K){20,20}
               (K){20}
     1:        KKKKKKKKKKKKKKKKKKKK KKVTK

               (K){20}
     2:      K KKKKKKKKKKKKKKKKKKKK KVTK

               (K){20}
     3:     KK KKKKKKKKKKKKKKKKKKKK VTK

    AAW38845  ck: 5054  len: 34   ! Aaw38845 Delivery peptide used in peptide m

1            (R,K){20,20}
               (K){20}
     1:        KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     2:      K KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     3:     KK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     4:    KKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     5:   KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     6:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     7:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     8:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     9:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

               (K){20}
    10:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

               (K){20}
    11:  KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

               (K){20}
    12:  KKKKK KKKKKKKKKKKKKKKKKKKK KXK

               (K){20}
    13:  KKKKK KKKKKKKKKKKKKKKKKKKK XK

    AAW38880  ck: 6814  len: 26   ! Aaw38880 Delivery peptide used in peptide m

1            (R,K){20,20}
               (K){20}
     1:        KKKKKKKKKKKKKKKKKKKK KKKVT

               (K){20}
     2:      K KKKKKKKKKKKKKKKKKKKK KKVTK

               (K){20}
     3:     KK KKKKKKKKKKKKKKKKKKKK KVTK

               (K){20}
     4:    KKK KKKKKKKKKKKKKKKKKKKK VTK

    AAW38846  ck: 7692  len: 35   ! Aaw38846 Delivery peptide used in peptide m

1            (R,K){20,20}
               (K){20}
     1:        KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     2:      K KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     3:     KK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     4:    KKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     5:   KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     6:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     7:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     8:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     9:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    10:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

               (K){20}
    11:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK
```

```
                 (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38833   ck: 9248   len: 22    ! Aaw38833 Delivery peptide used in peptide m

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK XK

AAW38834   ck: 986    len: 23    ! Aaw38834 Delivery peptide used in peptide m

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
 2:           K KKKKKKKKKKKKKKKKKKKK XK

AAW38835   ck: 2799   len: 24    ! Aaw38835 Delivery peptide used in peptide m

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
 2:           K KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
 3:          KK KKKKKKKKKKKKKKKKKKKK XK

AAW38836   ck: 4687   len: 25    ! Aaw38836 Delivery peptide used in peptide m

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
 2:           K KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
 3:          KK KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
 4:         KKK KKKKKKKKKKKKKKKKKKKK XK

AAW38837   ck: 6650   len: 26    ! Aaw38837 Delivery peptide used in peptide m

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
 2:           K KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
 3:          KK KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
 4:         KKK KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
```

```
 5:  KKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38838   ck: 8688   len: 27    ! Aaw38838 Delivery peptide used in peptide ma

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 2:           K KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
 3:          KK KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
 4:         KKK KKKKKKKKKKKKKKKKKKKK KKXK

                 (K){20}
 5:        KKKK KKKKKKKKKKKKKKKKKKKK KXK

                 (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38796   ck: 9227   len: 22    ! Aaw38796 Delivery peptide used in peptide ma

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK WK

AAW38797   ck: 964    len: 23    ! Aaw38797 Delivery peptide used in peptide ma

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK KWK

                 (K){20}
 2:           K KKKKKKKKKKKKKKKKKKKK WK

AAW38798   ck: 2776   len: 24    ! Aaw38798 Delivery peptide used in peptide ma

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK KKWK

                 (K){20}
 2:           K KKKKKKKKKKKKKKKKKKKK KWK

                 (K){20}
 3:          KK KKKKKKKKKKKKKKKKKKKK WK

AAW38799   ck: 4663   len: 25    ! Aaw38799 Delivery peptide used in peptide ma

1               (R,K){20,20}
                 (K){20}
 1:             KKKKKKKKKKKKKKKKKKKK KKKWK

                 (K){20}
 2:           K KKKKKKKKKKKKKKKKKKKK KKWK

                 (K){20}
 3:          KK KKKKKKKKKKKKKKKKKKKK KWK

                 (K){20}
 4:         KKK KKKKKKKKKKKKKKKKKKKK WK

AAW38800   ck: 6625   len: 26    ! Aaw38800 Delivery peptide used in peptide ma
```

```
1               (R,K){20,20}
                  (K){20}
        1:      KKKKKKKKKKKKKKKKKKKK KKKKW

                  (K){20}
        2:    K KKKKKKKKKKKKKKKKKKKK KKKWK

                  (K){20}
        3:   KK KKKKKKKKKKKKKKKKKKKK KKWK

                  (K){20}
        4:  KKK KKKKKKKKKKKKKKKKKKKK KWK

                  (K){20}
        5: KKKK KKKKKKKKKKKKKKKKKKKK WK

        AAW38801  ck: 8662  len: 27    ! Aaw38801 Delivery peptide used in peptide m
1               (R,K){20,20}
                  (K){20}
        1:      KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        2:    K KKKKKKKKKKKKKKKKKKKK KKKKW

                  (K){20}
        3:   KK KKKKKKKKKKKKKKKKKKKK KKKWK

                  (K){20}
        4:  KKK KKKKKKKKKKKKKKKKKKKK KKWK

                  (K){20}
        5: KKKK KKKKKKKKKKKKKKKKKKKK KWK

                  (K){20}
        6: KKKKK KKKKKKKKKKKKKKKKKKKK WK

        AAW38802  ck: 774   len: 28    ! Aaw38802 Delivery peptide used in peptide m
1               (R,K){20,20}
                  (K){20}
        1:      KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        2:    K KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        3:   KK KKKKKKKKKKKKKKKKKKKK KKKKW

                  (K){20}
        4:  KKK KKKKKKKKKKKKKKKKKKKK KKKWK

                  (K){20}
        5: KKKK KKKKKKKKKKKKKKKKKKKK KKWK

                  (K){20}
        6: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

                  (K){20}
        7: KKKKK KKKKKKKKKKKKKKKKKKKKK WK

        AAW38803  ck: 2961  len: 29    ! Aaw38803 Delivery peptide used in peptide m
1               (R,K){20,20}
                  (K){20}
        1:      KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        2:    K KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        3:   KK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        4:  KKK KKKKKKKKKKKKKKKKKKKK KKKKW

                  (K){20}
        5: KKKK KKKKKKKKKKKKKKKKKKKK KKKWK

                  (K){20}
        6: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

                  (K){20}
        7: KKKKK KKKKKKKKKKKKKKKKKKKKK KWK

                  (K){20}
        8: KKKKK KKKKKKKKKKKKKKKKKKKKKK WK

        AAW38804  ck: 5223  len: 30    ! Aaw38804 Delivery peptide used in peptide m.
1               (R,K){20,20}
                  (K){20}
        1:      KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        2:    K KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        3:   KK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        4:  KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        5: KKKK KKKKKKKKKKKKKKKKKKKK KKKKW

                  (K){20}
        6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

                  (K){20}
        7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKWK

                  (K){20}
        8: KKKKK KKKKKKKKKKKKKKKKKKKKKK KWK

                  (K){20}
        9: KKKKK KKKKKKKKKKKKKKKKKKKKKKK WK

        AAW38805  ck: 7560  len: 31    ! Aaw38805 Delivery peptide used in peptide m.
1               (R,K){20,20}
                  (K){20}
        1:      KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        2:    K KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        3:   KK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        4:  KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        5: KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

                  (K){20}
        7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKWK
```

```
             (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

             (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

             (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38806  ck: 9972  len: 32   ! Aaw38806 Delivery peptide used in peptide m

1           (R,K){20,20}
             (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

             (K){20}
 8:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

             (K){20}
 9:   KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

             (K){20}
10:   KKKKK KKKKKKKKKKKKKKKKKKKK KWK

             (K){20}
11:   KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38807  ck: 2459  len: 33   ! Aaw38807 Delivery peptide used in peptide m

1           (R,K){20,20}
             (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 8:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

             (K){20}
 9:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

             (K){20}
10:   KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

             (K){20}
11:   KKKKK KKKKKKKKKKKKKKKKKKKK KWK

             (K){20}
12:   KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38808  ck: 5021  len: 34   ! Aaw38808 Delivery peptide used in peptide ma

1           (R,K){20,20}
             (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 8:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 9:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

             (K){20}
10:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

             (K){20}
11:   KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

             (K){20}
12:   KKKKK KKKKKKKKKKKKKKKKKKKK KWK

             (K){20}
13:   KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38881  ck: 8859  len: 27   ! Aaw38881 Delivery peptide used in peptide ma

1           (R,K){20,20}
             (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKV

             (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKVT

             (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKVTK

             (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KVTK
```

```
              (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK VTK

  AAW38847  ck: 405   len: 36    ! Aaw38847 Delivery peptide used in peptide m

1           (R,K){20,20}
              (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
   12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

              (K){20}
   13: KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

              (K){20}
   14: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

              (K){20}
   15: KKKKK KKKKKKKKKKKKKKKKKKKK XK

  AAW38882  ck: 979   len: 28    ! Aaw38882 Delivery peptide used in peptide m

1           (R,K){20,20}
              (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKKV

              (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKKVT

              (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KKVTK

              (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK KVTK

              (K){20}
    6: KKKKK KKKKKKKKKKKKKKKKKKKK VTK
```

```
  AAW38848  ck: 3193  len: 37    ! Aaw38848 Delivery peptide used in peptide m

1           (R,K){20,20}
              (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
   13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

              (K){20}
   14: KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

              (K){20}
   15: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

              (K){20}
   16: KKKKK KKKKKKKKKKKKKKKKKKKK XK

  AAW38883  ck: 3174  len: 29    ! Aaw38883 Delivery peptide used in peptide m

1           (R,K){20,20}
              (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKKKV

              (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KKKVT

              (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK KKVTK

              (K){20}
```

```
   6: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

            (K){20}
   7: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

  AAW38849  ck: 6056  len: 38   ! Aaw38849 Delivery peptide used in peptide m

1           (R,K){20,20}
            (K){20}
   1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

            (K){20}
  16: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

            (K){20}
  17: KKKKK KKKKKKKKKKKKKKKKKKKK KK

  AAW38884  ck: 5444  len: 30   ! Aaw38884 Delivery peptide used in peptide m

1           (R,K){20,20}
            (K){20}
   1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKV

            (K){20}
   5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKVT

            (K){20}
   6: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

            (K){20}
   7: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

            (K){20}
   8: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

  AAW38850  ck: 8994  len: 39   ! Aaw38850 Delivery peptide used in peptide ma

1           (R,K){20,20}
            (K){20}
   1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
   9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
  15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

            (K){20}
  16: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

            (K){20}
  17: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

            (K){20}
  18: KKKKK KKKKKKKKKKKKKKKKKKKK KK
```

```
          AAW38885  ck: 7789  len: 31    ! Aaw38885 Delivery peptide used in peptide m
1                   (R,K){20,20}
                      (K){20}
           1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKV

                      (K){20}
           6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

                      (K){20}
           7: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

                      (K){20}
           8: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

                      (K){20}
           9: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

          AAW38851  ck: 2007  len: 40    ! Aaw38851 Delivery peptide used in peptide m
1                   (R,K){20,20}
                      (K){20}
           1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                      (K){20}
          16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                      (K){20}
          17: KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

                      (K){20}
          18: KKKKK KKKKKKKKKKKKKKKKKKKK KXK

                      (K){20}
          19: KKKKK KKKKKKKKKKKKKKKKKKKK XK

          AAW38886  ck: 209  len: 32    ! Aaw38886 Delivery peptide used in peptide m
1                   (R,K){20,20}
                      (K){20}
           1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

                      (K){20}
           7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

                      (K){20}
           8: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

                      (K){20}
           9: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

                      (K){20}
          10: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

          AAW38852  ck: 5095  len: 41    ! Aaw38852 Delivery peptide used in peptide m
1                   (R,K){20,20}
                      (K){20}
           1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
```

```
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKX

           (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKXK

           (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKXK

           (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KXK

           (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK XK

AAW38887  ck: 2704  len: 33    ! Aaw38887 Delivery peptide used in peptide m

1          (R,K){20,20}
           (K){20}
 1:        KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 2:      K KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 3:     KK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 4:    KKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 5:   KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKV

           (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKVT

           (K){20}
```

```
 9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKVTK

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KVTK

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK VTK

AAW38853  ck: 8258  len: 42    ! Aaw38853 Delivery peptide used in peptide ma

1          (R,K){20,20}
           (K){20}
 1:        KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 2:      K KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 3:     KK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 4:    KKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 5:   KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKX

           (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKXK

           (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKXK

           (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKKK KXK

           (K){20}
```

```
21: KKKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38888  ck: 5274  len: 34    ! Aaw38888 Delivery peptide used in peptide m

1               (R,K){20,20}
                   (K){20}
 1:           KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 2:         K KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 3:        KK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 4:       KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 5:      KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 6:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 7:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 8:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

                   (K){20}
 9:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

                   (K){20}
10:     KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

                   (K){20}
11:     KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

                   (K){20}
12:     KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38854  ck: 1496  len: 43    ! Aaw38854 Delivery peptide used in peptide m

1               (R,K){20,20}
                   (K){20}
 1:           KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 2:         K KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 3:        KK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 4:       KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 5:      KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 6:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 7:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 8:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 9:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
10:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
11:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
12:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
13:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
14:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
15:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
16:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
17:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
18:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                   (K){20}
19:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                   (K){20}
20:     KKKKK KKKKKKKKKKKKKKKKKKKK KKXK

                   (K){20}
21:     KKKKK KKKKKKKKKKKKKKKKKKKK KXK

                   (K){20}
22:     KKKKK KKKKKKKKKKKKKKKKKKKK XK

AAW38889  ck: 7919  len: 35    ! Aaw38889 Delivery peptide used in peptide m.

1               (R,K){20,20}
                   (K){20}
 1:           KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 2:         K KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 3:        KK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 4:       KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 5:      KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 6:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 7:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 8:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 9:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

                   (K){20}
10:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT
```

```
            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

            (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

            (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38890  ck: 639   len: 36    ! Aaw38890 Delivery peptide used in peptide m
1
          (R,K){20,20}
            (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

            (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

            (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

            (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38891  ck: 3434  len: 37    ! Aaw38891 Delivery peptide used in peptide m
1
          (R,K){20,20}
            (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

            (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

            (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

            (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

            (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38892  ck: 6304  len: 38    ! Aaw38892 Delivery peptide used in peptide ma
1
          (R,K){20,20}
            (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

            (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT
```

```
                 (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

                 (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

                 (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38893  ck: 9249  len: 39    ! Aaw38893 Delivery peptide used in peptide m

1              (R,K){20,20}
                 (K){20}
 1:           KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

                 (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

                 (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

                 (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

                 (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38894  ck: 2269  len: 40    ! Aaw38894 Delivery peptide used in peptide m

1              (R,K){20,20}
                 (K){20}
 1:           KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

                 (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

                 (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

                 (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

                 (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38895  ck: 5364  len: 41    ! Aaw38895 Delivery peptide used in peptide m

1              (R,K){20,20}
                 (K){20}
 1:           KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
```

```
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

            (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

            (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

            (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

            (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38896   ck: 8534   len: 42    ! Aaw38896 Delivery peptide used in peptide m

1          (R,K){20,20}
            (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

            (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT

            (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK

            (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK

            (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38897   ck: 1779   len: 43    ! Aaw38897 Delivery peptide used in peptide ma

1          (R,K){20,20}
            (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
```

```
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV
              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT
              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK
              (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK
              (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38898  ck: 5099  len: 44    ! Aaw38898 Delivery peptide used in peptide m

1           (R,K){20,20}
              (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 8:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 9:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
10:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
11:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
12:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
13:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
14:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
15:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}

16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV
              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVT
              (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKVTK
              (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KVTK
              (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK VTK

AAW38809  ck: 7658  len: 35    ! Aaw38809 Delivery peptide used in peptide m

1           (R,K){20,20}
              (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 8:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
 9:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
10:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW
              (K){20}
11:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK
              (K){20}
12:   KKKKK KKKKKKKKKKKKKKKKKKKK KKWK
              (K){20}
13:   KKKKK KKKKKKKKKKKKKKKKKKKK KWK
              (K){20}
14:   KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38810  ck: 370   len: 36    ! Aaw38810 Delivery peptide used in peptide m

1           (R,K){20,20}
              (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

          (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

          (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

          (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

          (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38811  ck: 3157  len: 37    ! Aaw38811 Delivery peptide used in peptide m
1         (R,K){20,20}
          (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

          (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

          (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

          (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

          (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38812  ck: 6019  len: 38    ! Aaw38812 Delivery peptide used in peptide m
1         (R,K){20,20}
          (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

          (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

          (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

          (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KWK
```

```
              (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38813  ck: 8956  len: 39    ! Aaw38813 Delivery peptide used in peptide m

1         (R,K){20,20}
              (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 8:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 9:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
10:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
11:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
12:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
13:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
14:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

              (K){20}
15:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

              (K){20}
16:   KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

              (K){20}
17:   KKKKK KKKKKKKKKKKKKKKKKKKK KWK

              (K){20}
18:   KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38814  ck: 1968  len: 40    ! Aaw38814 Delivery peptide used in peptide m

1         (R,K){20,20}
              (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 3:      KK KKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 8:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 9:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
10:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
11:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
12:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
13:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
14:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
15:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

              (K){20}
16:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

              (K){20}
17:   KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

              (K){20}
18:   KKKKK KKKKKKKKKKKKKKKKKKKK KWK

              (K){20}
19:   KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38815  ck: 5055  len: 41    ! Aaw38815 Delivery peptide used in peptide m

1         (R,K){20,20}
              (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
             (K){20}
  8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

             (K){20}
 17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

             (K){20}
 18: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

             (K){20}
 19: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

             (K){20}
 20: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAW38816  ck: 8217  len: 42    ! Aaw38816 Delivery peptide used in peptide m

1           (R,K){20,20}
             (K){20}
  1:       KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  3:    KK KKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
             (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 17: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKW

             (K){20}
 18: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKWK

             (K){20}
 19: KKKKK KKKKKKKKKKKKKKKKKKKKK KKWK

             (K){20}
 20: KKKKK KKKKKKKKKKKKKKKKKKKKK KWK

             (K){20}
 21: KKKKK KKKKKKKKKKKKKKKKKKKKK WK

AAW38817  ck: 1454  len: 43    ! Aaw38817 Delivery peptide used in peptide ma

1           (R,K){20,20}
             (K){20}
  1:       KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
    13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

                (K){20}
    19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

                (K){20}
    20: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

                (K){20}
    21: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

                (K){20}
    22: KKKKK KKKKKKKKKKKKKKKKKKKK WK

    AAW24865  ck: 2211  len: 40    ! Aaw24865 Bifunctional peptide I for binding

1             (R,K){20,20}
              · (K){20}
    21: YEDES KKKKKKKKKKKKKKKKKKKK

    AAW24450  ck: 8137  len: 45    ! Aaw24450 Nucleic acid (NA) binding peptide

1             (R,K){20,20}
                (K){20}
     4:   YKA KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     5:  YKAK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     6: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     7: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     8: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}

    14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

                (K){20}
    21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

                (K){20}
    22: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

                (K){20}
    23: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

                (K){20}
    24: KKKKK KKKKKKKKKKKKKKKKKKKK WK

    AAW21590  ck: 4875  len: 30    ! Aaw21590 Antibiotic potentiating peptide #2

1             (R,K){20,20}
                (K){20}
     1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
     8: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
     9: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
    10: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
    11: KKKKK KKKKKKKKKKKKKKKKKKKK

    AAW21591  ck: 5075  len: 434   ! Aaw21591 Antibiotic potentiating peptide #3

1             (R,K){20,20}
```

```
          (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
    (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
74: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
78: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
79: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
81: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
82: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
90: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

    (K){20}
```

```
 98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
 99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
108: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
109: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
110: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
111: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
112: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
113: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
115: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
116: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
117: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
118: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
121: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
122: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
123: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
124: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
125: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
126: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
127: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
128: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
129: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
130: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
131: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
132: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
133: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
134: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
135: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
136: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
137: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
138: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
139: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
140: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
141: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
142: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
143: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
144: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
145: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
               (K){20}
146: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
147: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
148: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
149: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
150: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
151: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
152: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
153: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
154: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
155: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
156: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
157: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
158: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
159: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
160: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
161: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
162: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
163: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
164: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
165: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
166: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
167: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
168: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
169: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
170: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
171: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
172: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
173: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
174: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
175: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
176: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
177: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
178: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
179: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
180: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
181: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
182: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
183: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
184: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
185: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
186: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
187: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
188: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
189: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
190: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
191: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
192: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
193: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
194: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
195: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                    (K){20}
196: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
197: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
198: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
199: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
200: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
201: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
202: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
203: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
204: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
205: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
206: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
207: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
208: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
209: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
210: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
211: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
212: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
213: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
214: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
215: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
216: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
217: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
218: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
219: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
220: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
221: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
222: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
223: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
224: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
225: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
226: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
227: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
228: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
229: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
230: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
231: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
232: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
233: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
234: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
235: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
236: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
237: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
238: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
239: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
240: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
241: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
242: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
243: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                    (K){20}
```

244: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
245: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
246: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
247: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
248: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
249: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
250: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
251: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
252: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
253: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
254: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
255: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
256: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
257: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
258: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
259: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
260: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
261: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
262: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
263: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
264: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
265: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
266: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
267: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
268: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
269: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
270: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
271: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
272: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
273: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
274: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
275: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
276: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
277: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
278: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
279: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
280: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
281: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
282: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
283: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
284: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
285: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
286: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
287: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
288: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
289: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
290: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
291: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
292: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
293: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
294: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
295: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
296: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
297: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
298: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
299: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
300: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
301: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
302: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
303: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
304: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
305: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
306: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
307: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
308: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
309: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
310: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
311: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
312: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
313: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
314: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
315: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
316: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
317: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
318: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
319: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
320: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
321: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
322: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
323: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
324: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
325: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
326: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
327: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
328: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
329: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
330: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
331: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
332: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
333: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
334: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
335: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
336: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
337: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
338: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
339: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
340: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

(K){20}
341: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

```
         (K){20}
342: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
343: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
344: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
345: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
346: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
347: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
348: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
349: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
350: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
351: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
352: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
353: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
354: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
355: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
356: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
357: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
358: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
359: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
360: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
361: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
362: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
363: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
364: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
365: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
366: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
367: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
368: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
369: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
370: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
371: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
372: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
373: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
374: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
375: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
376: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
377: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
378: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
379: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
380: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
381: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
382: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
383: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
384: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
385: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
386: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
387: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
388: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
389: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
         (K){20}
```

```
390: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
391: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
392: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
393: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
394: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
395: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
396: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
397: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
398: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
399: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
400: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
401: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
402: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
403: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
404: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
405: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
406: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
407: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
408: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
409: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
410: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
411: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK

               (K){20}
412: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
413: KKKKK KKKKKKKKKKKKKKKKKKKKK KK

               (K){20}
414: KKKKK KKKKKKKKKKKKKKKKKKKKK K

               (K){20}
415: KKKKK KKKKKKKKKKKKKKKKKKKKK

AAW65939  ck: 1569  len: 40   ! Aaw65939 Polylysine peptide NBC32. 11/1998

1             (R,K){20,20}
               (K){20}
 2:     T KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 3:    TK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 4:   TKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 5:  TKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 6: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKY

               (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKYC

               (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKYCG

               (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KYCG

               (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK YCG

AAW48808  ck: 9658  len: 56   ! Aaw48808 Homo sapiens clone CG109_1 protein

1             (R,K){20,20}
               (K){20}
34: EEFRE KKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
35: EFREK KKKKKKKKKKKKKKKKKKKK KK

               (K){20}
36: FREKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
```

```
        37: REKKK KKKKKKKKKKKKKKKKKKKK

        AAW45801  ck: 9500  len: 39   ! Aaw45801 One chain of a bombesin dimer. 6/1

1                  (R,K){20,20}
                     (K){20}
         2:      C KKKKKKKKKKKKKKKKKKKK XGGGQ

        AAW45800  ck: 9056  len: 39   ! Aaw45800 One chain of an alpha-melanocyte s

1                  (R,K){20,20}
                     (K){20}
        18: VGGGX KKKKKKKKKKKKKKKKKKKK CX

        AAW45802  ck: 7664  len: 35   ! Aaw45802 One chain of an alpha-MSH receptor

1                  (R,K){20,20}
                     (K){20}
        14: VGGGX KKKKKKKKKKKKKKKKKKKK CX

        AAY43246  ck: 9752  len: 32   ! Aay43246 Cell-surface molecule binding pept

1                  (R,K){20,20}
                     (K){20}
        12: SGSGS KKKKKKKKKKKKKKKKKKKK K

                     (K){20}
        13: GSGSK KKKKKKKKKKKKKKKKKKKK

        AAY07213  ck: 2211  len: 40   ! Aay07213 Peptide transfection vector #I. 7/

1                  (R,K){20,20}
                     (K){20}
        21: YEDES KKKKKKKKKKKKKKKKKKKK

        AAY12950  ck: 5821  len: 62   ! Aay12950 Amino acid sequence of a human sec

1                  (R,K){20,20}
                     (K){20}
        40: QFQAS KKKKKKKKKKKKKKKKKKKK KKK

                     (K){20}
        41: FQASK KKKKKKKKKKKKKKKKKKKK KK

                     (K){20}
        42: QASKK KKKKKKKKKKKKKKKKKKKK K

                     (K){20}
        43: ASKKK KKKKKKKKKKKKKKKKKKKK

        AAB59105  ck: 8456  len: 27   ! Aab59105 Breast and ovarian cancer associat

1                  (R,K){20,20}
                     (K){20}
         6: NSAXX KKKKKKKKKKKKKKKKKKKK KK

                     (K){20}
         7: SAXXK KKKKKKKKKKKKKKKKKKKK K

                     (K){20}
         8: AXXKK KKKKKKKKKKKKKKKKKKKK

        AAB53249  ck: 4945  len: 59   ! Aab53249 Human colon cancer antigen protein

1                  (R,K){20,20}
```

```
                     (K){20}
        27: ALLSL KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        28: LLSLK KKKKKKKKKKKKKKKKKKKK KKKKN

                     (K){20}
        29: LSLKK KKKKKKKKKKKKKKKKKKKK KKKNX

                     (K){20}
        30: SLKKK KKKKKKKKKKKKKKKKKKKK KKNXG

                     (K){20}
        31: LKKKK KKKKKKKKKKKKKKKKKKKK KNXGG

                     (K){20}
        32: KKKKK KKKKKKKKKKKKKKKKKKKK NXGGG

        AAB53659  ck: 3850  len: 184  ! Aab53659 Human colon cancer antigen protein

1                  (R,K){20,20}
                     (K){20}
         1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
         2:     K KKKKKKKKKKKKKKKKKKKK KKKKX

                     (K){20}
         3:    KK KKKKKKKKKKKKKKKKKKKK KKKXG

                     (K){20}
         4:   KKK KKKKKKKKKKKKKKKKKKKK KKXGG

                     (K){20}
         5:  KKKK KKKKKKKKKKKKKKKKKKKK KXGGR

                     (K){20}
         6: KKKKK KKKKKKKKKKKKKKKKKKKK XGGRF

        AAB53800  ck: 296   len: 69   ! Aab53800 Human colon cancer antigen protein

1                  (R,K){20,20}
                     (K){20}
        20: NVDCE KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        21: VDCEK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        22: DCEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        23: CEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        24: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                     (K){20}
        29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXG

                (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKXGG

                (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KXGGP

                (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK XGGPF

AAB53806   ck: 8373   len: 64     ! Aab53806 Human colon cancer antigen protein

1              (R,K){20,20}
                (K){20}
25: IDCDS KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
26: DCDSK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
27: CDSKK KKKKKKKKKKKKKKKKKKKK KKKXK

                (K){20}
28: DSKKK KKKKKKKKKKKKKKKKKKKK KKXKX

                (K){20}
29: SKKKK KKKKKKKKKKKKKKKKKKKK KXKXK

                (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK XKXKK

AAB53977   ck: 7756   len: 75     ! Aab53977 Human colon cancer antigen protein

1              (R,K){20,20}
                (K){20}
24: QFCLL KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
25: FCLLK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
26: CLLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
27: LLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
```

```
28: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

                (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KGGPV

                (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK GGPVX

AAB53980   ck: 881   len: 45     ! Aab53980 Human colon cancer antigen protein

1              (R,K){20,20}
                (K){20}
 8: ICSFQ KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
```

```
  9: CSFQK KKKKKKKKKKKKKKKKKKKK KKKXG

          (K){20}
 10: SFQKK KKKKKKKKKKKKKKKKKKKK KKXGG

          (K){20}
 11: FQKKK KKKKKKKKKKKKKKKKKKKK KXGGR

          (K){20}
 12: QKKKK KKKKKKKKKKKKKKKKKKKK XGGRF

AAB54314   ck: 6235   len: 55    ! Aab54314 Human pancreatic cancer antigen pr

1        (R,K){20,20}
          (K){20}
 24: FFXXX KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 25: FXXXK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 26: XXXKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 27: XXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 28: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

          (K){20}
 29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

          (K){20}
 30: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGR

          (K){20}
 31: KKKKK KKKKKKKKKKKKKKKKKKKK KGGRF

          (K){20}
 32: KKKKK KKKKKKKKKKKKKKKKKKKK GGRF

AAB56121   ck: 5941   len: 125   ! Aab56121 Human secreted protein sequence en

1        (R,K){20,20}
          (K){20}
 87: ILKSQ KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 88: LKSQK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 89: KSQKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 90: SQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 91: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 92: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

          (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

          (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

          (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KGGP

          (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK GGP

AAB27956   ck: 6732   len: 139   ! Aab27956 Human secreted protein SEQ ID NO:

1        (R,K){20,20}
          (K){20}
 91: HFRPG KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 92: FRPGK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 93: RPGKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 94: PGKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 95: GKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
108: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
109: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
110: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
111: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
112: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
113: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
115: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
116: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
117: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
118: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKK

 AAB39140  ck: 9531  len: 66    ! Aab39140 Human secreted protein #48. 2/2001

1           (R,K){20,20}
              (K){20}
 41: MWTVX KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 42: WTVXK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 43: TVXKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
 44: VXKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
 45: XKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
 46: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
 47: KKKKK KKKKKKKKKKKKKKKKKKKK

 AAB41457  ck: 4536  len: 168   ! Aab41457 Human ORFX ORF1221 polypeptide seq

1           (R,K){20,20}
              (K){20}
144: KQQKP KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
145: QQKPK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
146: QKPKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
147: KPKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
148: PKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
149: KKKKK KKKKKKKKKKKKKKKKKKKK

 AAB42786  ck: 4959  len: 102   ! Aab42786 Human ORFX ORF2550 polypeptide sequ

1           (R,K){20,20}
              (K){20}
 81: SDVLQ KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
 82: DVLQK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
 83: VLQKK KKKKKKKKKKKKKKKKKKKK

 AAB42889  ck: 6432  len: 62    ! Aab42889 Human ORFX ORF2653 polypeptide sequ

1           (R,K){20,20}
              (K){20}
 40: SHRCL KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
 41: HRCLK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
 42: RCLKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
 43: CLKKK KKKKKKKKKKKKKKKKKKKK

 AAB43641  ck: 5216  len: 133   ! Aab43641 Human cancer associated protein sec

1           (R,K){20,20}
              (K){20}
114: GHEQS KKKKKKKKKKKKKKKKKKKK

 AAB43835  ck: 4025  len: 223   ! Aab43835 Human cancer associated protein sec

1           (R,K){20,20}
              (K){20}
196: NILFW KKKKKKKKKKKKKKKKKKKK XXGGA

 AAB44188  ck: 4991  len: 43    ! Aab44188 Human cancer associated protein sec

1           (R,K){20,20}
              (K){20}
  2:     E KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  3:    EK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
```

```
        4:  EKK KKKKKKKKKKKKKKKKKKKK KKKGR

                (K){20}
        5:  EKKK KKKKKKKKKKKKKKKKKKKK KKGRP

                (K){20}
        6: EKKKK KKKKKKKKKKKKKKKKKKKK KGRPF

                (K){20}
        7: KKKKK KKKKKKKKKKKKKKKKKKKK GRPFX

       AAB44380  ck: 8490  len: 42    ! Aab44380 Human secreted protein encoded by

1              (R,K){20,20}
                (K){20}
       22: IWKKI KKKKKKKKKKKKKKKKKKKK K

                (K){20}
       23: WKKIK KKKKKKKKKKKKKKKKKKKK

       AAB34331  ck: 8131  len: 66    ! Aab34331 Human secreted protein sequence en

1              (R,K){20,20}
                (K){20}
       40: IHHKP KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       41: HHKPK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       42: HKPKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       43: KPKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
       44: PKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
       45: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
       46: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
       47: KKKKK KKKKKKKKKKKKKKKKKKKK

       AAB23585  ck: 9893  len: 36    ! Aab23585 AsK21 linker peptide. 1/2001

1              (R,K){20,20}
                (K){20}
       15: PGSGS KKKKKKKKKKKKKKKKKKKK KG

                (K){20}
       16: GSGSK KKKKKKKKKKKKKKKKKKKK G

       AAB23586  ck: 58    len: 36    ! Aab23586 GsK21 linker peptide. 1/2001

1              (R,K){20,20}
                (K){20}
       15: SGSGS KKKKKKKKKKKKKKKKKKKK KG

                (K){20}
       16: GSGSK KKKKKKKKKKKKKKKKKKKK G

       AAB23591  ck: 7684  len: 630   ! Aab23591 Modified fibre protein encoded in

1              (R,K){20,20}

                (K){20}
      596: PGSGS KKKKKKKKKKKKKKKKKKKK KGSYS

                (K){20}
      597: GSGSK KKKKKKKKKKKKKKKKKKKK GSYSM

       AAB23592  ck: 7647  len: 630   ! Aab23592 Modified fibre protein encoded in

1              (R,K){20,20}
                (K){20}
      596: SGSGS KKKKKKKKKKKKKKKKKKKK KGSYS

                (K){20}
      597: GSGSK KKKKKKKKKKKKKKKKKKKK GSYSM

       AAB23593  ck: 2249  len: 640   ! Aab23593 Modified fibre protein encoded in

1              (R,K){20,20}
                (K){20}
      596: PGSGS KKKKKKKKKKKKKKKKKKKK KGSAE

                (K){20}
      597: GSGSK KKKKKKKKKKKKKKKKKKKK GSAEK

       AAB23594  ck: 2612  len: 640   ! Aab23594 Modified fibre protein encoded in

1              (R,K){20,20}
                (K){20}
      596: SGSGS KKKKKKKKKKKKKKKKKKKK KGSAE

                (K){20}
      597: GSGSK KKKKKKKKKKKKKKKKKKKK GSAEK

       AAB13780  ck: 7317  len: 21    ! Aab13780 Soluble peptide antigen pK. 11/200

1              (R,K){20,20}
                (K){20}
        2:     C KKKKKKKKKKKKKKKKKKKK

       AAB13783  ck: 4553  len: 45    ! Aab13783 Soluble tandem pEA/ pK peptide con

1              (R,K){20,20}
                (K){20}
       26: AAAAA KKKKKKKKKKKKKKKKKKKK

       AAB13784  ck: 4126  len: 44    ! Aab13784 Soluble tandem HA/ pK peptide conj

1              (R,K){20,20}
                (K){20}
       25: DGWYG KKKKKKKKKKKKKKKKKKKK

       AAG00834  ck: 6330  len: 103   ! Aag00834 Human secreted protein, SEQ ID NO:

1              (R,K){20,20}
                (K){20}
       83: RKRQQ KKKKKKKKKKKKKKKKKKKK K

                (K){20}
       84: KRQQK KKKKKKKKKKKKKKKKKKKK

       AAY98493  ck: 8137  len: 45    ! Aay98493 Peptide #5 used in nucleic acid tr

1              (R,K){20,20}
                (K){20}
        4:   YKA KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
           (K){20}
 5:  YKAK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 6: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 8: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

           (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

           (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

           (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

           (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAY98495   ck: 4361   len: 59    ! Aay98495 Nuclear ligand used in nucleic aci

1          (R,K){20,20}
           (K){20}
18: APYKA KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
19: PYKAK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
20: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
           (K){20}
21: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
22: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

           (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

           (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

           (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

           (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAY98497   ck: 4925   len: 100   ! Aay98497 Peptide #8 used in nucleic acid tra

1          (R,K){20,20}
           (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
```

```
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
74: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
78: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
79: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
81: KKKKK KKKKKKKKKKKKKKKKKKKK

AAY81805  ck: 1885  len: 351   ! Aay81805 Murine mahogany protein sequence ak

1             (R,K){20,20}
               (R,K){20}
 3:    EF RKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 4:   EFR KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 5:  EFRK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 6: EFRKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 7: FRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 8: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
   23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
   24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
   25: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGR

                (K){20}
   26: KKKKK KKKKKKKKKKKKKKKKKKKK KGGRR

                (K){20}
   27: KKKKK KKKKKKKKKKKKKKKKKKKK GGRRG

   AAY56902  ck: 4875  len: 30   ! Aay56902 (Lys)30 protein sequence. 4/2000

1              (R,K){20,20}
                (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
    8: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
    9: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
   10: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
   11: KKKKK KKKKKKKKKKKKKKKKKKKK

   AAY56903  ck: 5075  len: 434   ! Aay56903 (Lys)434 protein sequence. 4/2000

1              (R,K){20,20}
                (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
        (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
74: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
78: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
```

```
 79: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 80: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 81: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 82: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 83: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 84: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 85: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 86: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 87: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 88: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 89: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 90: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 91: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 92: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 93: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 94: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 95: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 96: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 97: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 98: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 99: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
108: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
109: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
110: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
111: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
112: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
113: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
115: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
116: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
117: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
118: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
121: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
122: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
123: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
124: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
125: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
126: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
127: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
```

```
      (K){20}
128: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
129: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
130: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
131: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
132: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
133: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
134: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
135: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
136: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
137: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
138: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
139: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
140: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
141: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
142: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
143: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
144: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
145: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
146: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
147: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
148: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
149: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
150: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
151: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
152: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
153: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
154: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
155: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
156: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
157: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
158: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
159: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
160: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
161: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
162: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
163: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
164: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
165: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
166: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
167: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
168: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
169: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
170: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
171: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
172: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
173: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
174: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
175: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
      (K){20}
176: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
177: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
178: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
179: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
180: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
181: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
182: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
183: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
184: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
185: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
186: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
187: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
188: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
189: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
190: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
191: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
192: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
193: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
194: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
195: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
196: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
197: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
198: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
199: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
200: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
201: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
202: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
203: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
204: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
205: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
206: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
207: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
208: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
209: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
210: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
211: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
212: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
213: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
214: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
215: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
216: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
217: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
218: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
219: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
220: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
221: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
222: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
223: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
224: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
```

```
225: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
226: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
227: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
228: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
229: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
230: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
231: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
232: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
233: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
234: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
235: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
236: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
237: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
238: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
239: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
240: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
241: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
242: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
243: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
244: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
245: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
246: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
247: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
248: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
249: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
250: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
251: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
252: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
253: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
254: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
255: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
256: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
257: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
258: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
259: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
260: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
261: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
262: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
263: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
264: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
265: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
266: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
267: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
268: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
269: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
270: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
271: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
272: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
273: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
274: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
275: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
276: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
277: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
278: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
279: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
280: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
281: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
282: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
283: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
284: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
285: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
286: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
287: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
288: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
289: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
290: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
291: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
292: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
293: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
294: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
295: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
296: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
297: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
298: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
299: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
300: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
301: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
302: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
303: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
304: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
305: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
306: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
307: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
308: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
309: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
310: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
311: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
312: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
313: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
314: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
315: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
316: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
317: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
318: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
319: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
320: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
321: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
322: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
       (K){20}
323: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
324: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
325: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
326: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
327: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
328: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
329: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
330: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
331: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
332: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
333: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
334: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
335: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
336: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
337: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
338: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
339: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
340: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
341: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
342: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
343: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
344: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
345: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
346: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
347: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
348: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
349: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
350: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
351: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
352: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
353: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
354: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
355: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
356: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
357: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
358: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
359: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
360: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
361: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
362: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
363: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
364: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
365: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
366: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
367: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
368: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
369: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
370: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
       (K){20}
```

```
371: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
372: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
373: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
374: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
375: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
376: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
377: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
378: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
379: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
380: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
381: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
382: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
383: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
384: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
385: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
386: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
387: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
388: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
389: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
390: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
391: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
392: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
393: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
394: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
395: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
396: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
397: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
398: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
399: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
400: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
401: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
402: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
403: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
404: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
405: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
406: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
407: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
408: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
409: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
410: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
411: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

           (K){20}
412: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

           (K){20}
413: KKKKK KKKKKKKKKKKKKKKKKKKK KK

           (K){20}
414: KKKKK KKKKKKKKKKKKKKKKKKKK K

           (K){20}
415: KKKKK KKKKKKKKKKKKKKKKKKKK

AAY86248  ck: 8783  len: 128   ! Aay86248 Human secreted protein HCHPF68, SE(

1          (R,K){20,20}
           (K){20}
84: ANPPP KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
85: NPPPK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
86: PPPKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
 87: PPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 88: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 89: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 90: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 91: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 92: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

          (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

          (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

          (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKK KGGP

          (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKK GGP

  AAY59038  ck: 8137  len: 45    ! Aay59038 Peptide used in the construction c

1        (R,K){20,20}
          (K){20}
  4:   YKA KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
  5:  YKAK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
  6: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
  7: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
  8: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
  9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

          (K){20}
 21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

          (K){20}
 22: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

          (K){20}
 23: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

          (K){20}
 24: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAY59040  ck: 4361  len: 59    ! Aay59040 Nuclear ligand for transporting nu

1        (R,K){20,20}
          (K){20}
 18: APYKA KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 19: PYKAK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 20: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 21: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 22: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
           (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

           (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

           (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

           (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

           (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAY59042  ck: 4925  len: 100   ! Aay59042 Amino acid polymer seq ID NO: 62 c

          (R,K){20,20}
           (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
        (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
74: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
        (K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
        (K){20}
78: KKKKK KKKKKKKKKKKKKKKKKKKK KKK
        (K){20}
79: KKKKK KKKKKKKKKKKKKKKKKKKK KK
        (K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKKK K
```

```
              (K){20}
 81: KKKKK KKKKKKKKKKKKKKKKKKKK

 ABR01713   ck: 2112  len: 219   ! Abr01713 Human breast specific polypeptide

1           (R,K){20,20}
              (K){20}
190: PPPPP KKKKKKKKKKKKKKKKKKKK KKGGG

              (K){20}
191: PPPPK KKKKKKKKKKKKKKKKKKKK KGGGT

              (K){20}
192: PPPKK KKKKKKKKKKKKKKKKKKKK GGGTS

 AAE28374   ck: 5750  len: 20    ! Aae28374 Peptide #1 used for transfection e

1           (R,K){20,20}
              (K){20}
  1:       KKKKKKKKKKKKKKKKKKKK

 AAE28375   ck: 7220  len: 20    ! Aae28375 Peptide #1 used in the invention.

1           (R,K){20,20}
              (R){20}
  1:       RRRRRRRRRRRRRRRRRRRR

 AAE28376   ck: 4440  len: 40    ! Aae28376 Peptide #2 used in the invention.

1           (R,K){20,20}
             (R,K){20}
  1:       KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
  2:     K RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
  3:    KR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
  4:   KRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
  5:  KRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
  6: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
  7: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
  8: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
  9: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 10: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 11: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 12: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 13: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 14: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 15: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 16: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 17: RKRKR KRKRKRKRKRKRKRKRKRKR KRKR

             (R,K){20}
 18: KRKRK RKRKRKRKRKRKRKRKRKRK RKR

             (R,K){20}
 19: RKRKR KRKRKRKRKRKRKRKRKRKR KR

             (R,K){20}
 20: KRKRK RKRKRKRKRKRKRKRKRKRK R

             (R,K){20}
 21: RKRKR KRKRKRKRKRKRKRKRKRKR

 AAE28379   ck: 598   len: 23    ! Aae28379 RGD peptide #1 used in the inventi

1           (R,K){20,20}
              (K){20}
  1:       KKKKKKKKKKKKKKKKKKKK RGD

             (R,K){20}
  2:     K KKKKKKKKKKKKKKKKKKKR GD

 ABP66694   ck: 1944  len: 122   ! Abp66694 Human breast specific protein SEQ

1           (R,K){20,20}
              (K){20}
  8: AARAG KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  9: ARAGK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 10: RAGKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 11: AGKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 12: GKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
```

```
     19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                 (K){20}
     21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                 (K){20}
     22: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                 (K){20}
     23: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGF

                 (K){20}
     24: KKKKK KKKKKKKKKKKKKKKKKKKK GGGFV

     ABP66883  ck: 3983  len: 292   ! Abp66883 Human polypeptide SEQ ID NO 604. 1

1                (R,K){20,20}
                  (R,K){20}
    273: QVFAP RKKKKKKKKKKKKKKKKKKK

     ABP67072  ck: 7611  len: 315   ! Abp67072 Human polypeptide SEQ ID NO 793. 1

1                (R,K){20,20}
                  (R,K){20}
    273: QVFAP RKKKKKKKKKKKKKKKKKKK KGGRS

                 (K){20}
    274: VFAPR KKKKKKKKKKKKKKKKKKKK GGRSR

     ABG92583  ck: 7907  len: 39    ! Abg92583 Human DNA-binding protein #9. 11/2

1                (R,K){20,20}
                  (K){20}
      9: YFEDL KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     10: FEDLK KKKKKKKKKKKKKKKKKKKK KKKKG

                 (K){20}
     11: EDLKK KKKKKKKKKKKKKKKKKKKK KKKGX

                 (K){20}
     12: DLKKK KKKKKKKKKKKKKKKKKKKK KKGXA

                 (K){20}
     13: LKKKK KKKKKKKKKKKKKKKKKKKK KGXAA

                 (K){20}
     14: KKKKK KKKKKKKKKKKKKKKKKKKK GXAAL

     ABG92588  ck: 9194  len: 87    ! Abg92588 Human DNA-binding protein #14. 11/

1                (R,K){20,20}
                  (K){20}
     52: KIILL KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     53: IILLK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
     54: ILLKK KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
     55: LLKKK KKKKKKKKKKKKKKKKKKKK KKXKK

                 (K){20}
     56: LKKKK KKKKKKKKKKKKKKKKKKKK KXKKK

                 (K){20}
     57: KKKKK KKKKKKKKKKKKKKKKKKKK XKKKK

     ABG92589  ck: 8659  len: 104   ! Abg92589 Human DNA-binding protein #15. 11/

1                (R,K){20,20}
                  (K){20}
     75: PLGGQ KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     76: LGGQK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
     77: GGQKK KKKKKKKKKKKKKKKKKKKK KKKXG

                 (K){20}
     78: GQKKK KKKKKKKKKKKKKKKKKKKK KKXGX

                 (K){20}
     79: QKKKK KKKKKKKKKKKKKKKKKKKK KXGXP

                 (K){20}
     80: KKKKK KKKKKKKKKKKKKKKKKKKK XGXPF

     ABG92592  ck: 9398  len: 48    ! Abg92592 Human DNA-binding protein #18. 11/

1                (R,K){20,20}
                  (K){20}
      2:     Q KKKKKKKKKKKKKKKKKKKK KKGGR

                 (K){20}
      3:    QK KKKKKKKKKKKKKKKKKKKK KGGRX

                 (K){20}
      4:   QKK KKKKKKKKKKKKKKKKKKKK GGRXR

     ABG92598  ck: 8278  len: 53    ! Abg92598 Human DNA-binding protein #24. 11/

1                (R,K){20,20}
                  (K){20}
     30: NCGIL KKKKKKKKKKKKKKKKKKKK KKKK

                 (K){20}
     31: CGILK KKKKKKKKKKKKKKKKKKKK KKK

                 (K){20}
     32: GILKK KKKKKKKKKKKKKKKKKKKK KK

                 (K){20}
     33: ILKKK KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
     34: LKKKK KKKKKKKKKKKKKKKKKKKK

     ABG92599  ck: 444   len: 66    ! Abg92599 Human DNA-binding protein #25. 11/

1                (R,K){20,20}
                  (K){20}
     35: SMTFS KKKKKKKKKKKKKKKKKKKK KXGKK

                 (K){20}
     36: MTFSK KKKKKKKKKKKKKKKKKKKK XGKKK

     ABG92600  ck: 5503  len: 50    ! Abg92600 Human DNA-binding protein #26. 11/

1                (R,K){20,20}
```

```
                (K){20}
       30: IICLL KKKKKKKKKKKKKKKKKKKK K

                (K){20}
       31: ICLLK KKKKKKKKKKKKKKKKKKKK

       ABG92605  ck: 5691  len: 108   ! Abg92605 Human DNA-binding protein #31. 11/

1              (R,K){20,20}
                (K){20}
       78: VRPCL KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       79: RPCLK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       80: PCLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       81: CLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
       86: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
       87: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
       88: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
       89: KKKKK KKKKKKKKKKKKKKKKKKKK

       ABG92613  ck: 6029  len: 63    ! Abg92613 Human DNA-binding protein #39. 11/

1              (R,K){20,20}
                (K){20}
       40: KLTLL KKKKKKKKKKKKKKKKKKKK ISWG

       ABG92621  ck: 7170  len: 63    ! Abg92621 Human DNA-binding protein #47. 11/

1              (R,K){20,20}
                (K){20}
       37: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       38: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       39: SRAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       40: RAKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
       41: AKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
       42: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
       43: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
       44: KKKKK KKKKKKKKKKKKKKKKKKKK

       ABG92625  ck: 6110  len: 61    ! Abg92625 Human DNA-binding protein #51. 11/

1              (R,K){20,20}
                (K){20}
       28: RPTRP KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       29: PTRPK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       30: TRPKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       31: RPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       32: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
       34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
       35: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                (K){20}
       36: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGG

                (K){20}
       37: KKKKK KKKKKKKKKKKKKKKKKKKK GGGGG

       ABG92626  ck: 5764  len: 74    ! Abg92626 Human DNA-binding protein #52. 11/

1              (R,K){20,20}
                (K){20}
       40: EFLSA KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       41: FLSAK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       42: LSAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       43: SAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       44: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                   (K){20}
         49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                   (K){20}
         51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKX

                   (K){20}
         52: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

                   (K){20}
         53: KKKKK KKKKKKKKKKKKKKKKKKKK KX

                   (K){20}
         54: KKKKK KKKKKKKKKKKKKKKKKKKK X

         ABG92627  ck: 9217  len: 68    ! Abg92627 Human DNA-binding protein #53. 11/

1                 (R,K){20,20}
                   (K){20}
         38: FLFPE KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         39: LFPEK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         40: FPEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         41: PEKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                   (K){20}
         42: EKKKK KKKKKKKKKKKKKKKKKKKK KKKGX

                   (K){20}
         43: KKKKK KKKKKKKKKKKKKKKKKKKK KKGXX

                   (K){20}
         44: KKKKK KKKKKKKKKKKKKKKKKKKK KGXXF

                   (K){20}
         45: KKKKK KKKKKKKKKKKKKKKKKKKK GXXF

         ABG92629  ck: 8162  len: 79    ! Abg92629 Human DNA-binding protein #55. 11/

1                 (R,K){20,20}
                   (R,K){20}
         41: VRPRV RKKKKKKKKKKKKKKKKKKK KKKKG

                   (K){20}
         42: RPRVR KKKKKKKKKKKKKKKKKKKK KKKGG

                   (K){20}
         43: PRVRK KKKKKKKKKKKKKKKKKKKK KKGGR

                   (K){20}
         44: RVRKK KKKKKKKKKKKKKKKKKKKK KGGRF

                   (K){20}
         45: VRKKK KKKKKKKKKKKKKKKKKKKK GGRFR

         ABG92658  ck: 285   len: 118   ! Abg92658 Human DNA-binding protein #84. 11/

1                 (R,K){20,20}
                   (R,K){20}
         98: EKHKQ KKKKKKKKKKKKKKKKKKKR G
```

```
         ABG92659  ck: 5509  len: 58    ! Abg92659 Human DNA-binding protein #85. 11/

1                 (R,K){20,20}
                   (K){20}
         36: FYFVC KKKKKKKKKKKKKKKKKKKK KKK

                   (K){20}
         37: YFVCK KKKKKKKKKKKKKKKKKKKK KK

                   (K){20}
         38: FVCKK KKKKKKKKKKKKKKKKKKKK K

                   (K){20}
         39: VCKKK KKKKKKKKKKKKKKKKKKKK

         ABG92660  ck: 9074  len: 66    ! Abg92660 Human DNA-binding protein #86. 11/

1                 (R,K){20,20}
                   (K){20}
         40: LVQCE KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         41: VQCEK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         42: QCEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         43: CEKKK KKKKKKKKKKKKKKKKKKKK KKKK

                   (K){20}
         44: EKKKK KKKKKKKKKKKKKKKKKKKK KKK

                   (K){20}
         45: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                   (K){20}
         46: KKKKK KKKKKKKKKKKKKKKKKKKK K

                   (K){20}
         47: KKKKK KKKKKKKKKKKKKKKKKKKK

         ABG92661  ck: 8528  len: 150   ! Abg92661 Human DNA-binding protein #87. 11/

1                 (R,K){20,20}
                   (K){20}
        113: SRNTV KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        114: RNTVK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        115: NTVKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        116: TVKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        117: VKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        118: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                   (K){20}
        120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXX

                   (K){20}
        121: KKKKK KKKKKKKKKKKKKKKKKKKK KKXXA
```

```
                 (K){20}
      122: KKKKK KKKKKKKKKKKKKKKKKKKK KXXAV

                 (K){20}
      123: KKKKK KKKKKKKKKKKKKKKKKKKK XXAVL

       ABG92662  ck: 7676  len: 156   ! Abg92662 Human DNA-binding protein #88. 11/
1               (R,K){20,20}
                 (K){20}
      108: KTTWI KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      109: TTWIK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      110: TWIKK KKKKKKKKKKKKKKKKKKKK KKKKR

                 (K){20}
      111: WIKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                 (K){20}
      112: IKKKK KKKKKKKKKKKKKKKKKKKK KKRGX

                 (K){20}
      113: KKKKK KKKKKKKKKKKKKKKKKKKK KRGXG

                 (K){20}
      114: KKKKK KKKKKKKKKKKKKKKKKKKK RGXGS

                (R,K){20}
      115: KKKKK KKKKKKKKKKKKKKKKKKKR GXGST

       ABG92663  ck: 1736  len: 40    ! Abg92663 Human DNA-binding protein #89. 11/
1               (R,K){20,20}
                 (K){20}
       18: LPGSL KKKKKKKKKKKKKKKKKKKK KKK

                 (K){20}
       19: PGSLK KKKKKKKKKKKKKKKKKKKK KK

                 (K){20}
       20: GSLKK KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
       21: SLKKK KKKKKKKKKKKKKKKKKKKK

       ABG92665  ck: 1109  len: 98    ! Abg92665 Human DNA-binding protein #91. 11/
1               (R,K){20,20}
                 (K){20}
       53: QTKNT KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       54: TKNTK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       55: KNTKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       56: NTKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       57: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                 (K){20}
       65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                 (K){20}
       66: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGR

                 (K){20}
       67: KKKKK KKKKKKKKKKKKKKKKKKKK KGGRS

                 (K){20}
       68: KKKKK KKKKKKKKKKKKKKKKKKKK GGRSR

       ABG92667  ck: 8102  len: 111   ! Abg92667 Human DNA-binding protein #93. 11/
1               (R,K){20,20}
                 (K){20}
       78: EFHIL KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       79: FHILK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       80: HILKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       81: ILKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       87: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                 (K){20}
       89: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                 (K){20}
       90: KKKKK KKKKKKKKKKKKKKKKKKKK KK
```

```
            (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKKK K

            (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKK

ABG92668  ck: 8102  len: 111   ! Abg92668 Human DNA-binding protein #94. 11/

1         (R,K){20,20}
            (K){20}
78: EFHIL KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
79: FHILK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
80: HILKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
81: ILKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

            (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

            (K){20}
90: KKKKK KKKKKKKKKKKKKKKKKKKK KK

            (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKKK K

            (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKK

ABG92669  ck: 8319  len: 53   ! Abg92669 Human DNA-binding protein #95. 11/

1         (R,K){20,20}
            (K){20}
13: RYFKP KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
14: YFKPK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
15: FKPKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
16: KPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
17: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

            (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXE

            (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKXEG

            (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KXEGX

            (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK XEGX

ABG92671  ck: 7918  len: 80   ! Abg92671 Human DNA-binding protein #97. 11/2

1         (R,K){20,20}
            (K){20}
50: NVLTV KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
51: VLTVK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
52: LTVKK KKKKKKKKKKKKKKKKKKKK KKKKX

            (K){20}
53: TVKKK KKKKKKKKKKKKKKKKKKKK KKKXX

            (K){20}
54: VKKKK KKKKKKKKKKKKKKKKKKKK KKXXK

            (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KXXKK

            (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK XXKKK

ABG92673  ck: 4882  len: 41   ! Abg92673 Human DNA-binding protein #99. 11/2

1         (R,K){20,20}
            (K){20}
 8: FYCFF KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
```

```
 9: YCFFK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
10: CFFKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
11: FFKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
12: FKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG
          (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGX
          (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKGX
          (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KGX
          (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK GX

ABG92674  ck: 5469  len: 63    ! Abg92674 Human DNA-binding protein #100. 11

1        (R,K){20,20}
          (K){20}
30: IICLL KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
31: ICLLK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
32: CLLKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
33: LLKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX
          (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXG
          (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKXG
          (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KXG
          (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK XG

ABG92675  ck: 5075  len: 52    ! Abg92675 Human DNA-binding protein #101. 11

1        (R,K){20,20}
          (K){20}
30: FIVVX KKKKKKKKKKKKKKKKKKKK KKK
          (K){20}
31: IVVXK KKKKKKKKKKKKKKKKKKKK KK
          (K){20}
32: VVXKK KKKKKKKKKKKKKKKKKKKK K
          (K){20}
33: VXKKK KKKKKKKKKKKKKKKKKKKK

ABG92676  ck: 5741  len: 47    ! Abg92676 Human DNA-binding protein #102. 11

1        (R,K){20,20}
          (K){20}
20: ILTTF KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
21: LTTFK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
22: TTFKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
23: TFKKK KKKKKKKKKKKKKKKKKKKK KKKKX
          (K){20}
24: FKKKK KKKKKKKKKKKKKKKKKKKK KKKX
          (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKX
          (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KX
          (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK X

ABG92677  ck: 2868  len: 84    ! Abg92677 Human DNA-binding protein #103. 11

1        (R,K){20,20}
          (K){20}
53: KCTYE KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
54: CTYEK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
55: TYEKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
56: YEKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
57: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX
          (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK
```

```
               (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KXKK

               (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK XKK

ABG92678  ck: 4686  len: 73   ! Abg92678 Human DNA-binding protein #104. 11

1            (R,K){20,20}
               (K){20}
41: YLKKE KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
42: LKKEK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
43: KKEKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
44: KEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
45: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

               (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KK

               (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK

ABG92679  ck: 6676  len: 74   ! Abg92679 Human DNA-binding protein #105. 11

1            (R,K){20,20}
               (K){20}
47: LRTFQ KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
48: RTFQK KKKKKKKKKKKKKKKKKKKK KKKKX

               (K){20}
49: TFQKK KKKKKKKKKKKKKKKKKKKK KKKXX

               (K){20}
50: FQKKK KKKKKKKKKKKKKKKKKKKK KKXXG

               (K){20}
51: QKKKK KKKKKKKKKKKKKKKKKKKK KXXG

               (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK XXG

ABG92680  ck: 2283  len: 54   ! Abg92680 Human DNA-binding protein #106. 11/

1            (R,K){20,20}
               (K){20}
32: IVFCF KKKKKKKKKKKKKKKKKKKK KKX

               (K){20}
33: VFCFK KKKKKKKKKKKKKKKKKKKK KX

               (K){20}
34: FCFKK KKKKKKKKKKKKKKKKKKKK X

ABG92681  ck: 7503  len: 74   ! Abg92681 Human DNA-binding protein #107. 11/

1            (R,K){20,20}
               (K){20}
45: SHLTD KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
46: HLTDK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
47: LTDKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
48: TDKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
49: DKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

               (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KK

               (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK

ABG92683  ck: 5199  len: 84   ! Abg92683 Human DNA-binding protein #109. 11/

1            (R,K){20,20}
               (K){20}
63: AWNAS KKKKKKKKKKKKKKKKKKKK XG

ABG92684  ck: 7578  len: 31   ! Abg92684 Human DNA-binding protein #110. 11/

1            (R,K){20,20}
               (K){20}
 6: LTELE KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 7: TELEK KKKKKKKKKKKKKKKKKKKK KKKKX

               (K){20}
```

```
           (K){20}
8: ELEKK KKKKKKKKKKKKKKKKKKKK KKKX

           (K){20}
9: LEKKK KKKKKKKKKKKKKKKKKKKK KKX

           (K){20}
10: EKKKK KKKKKKKKKKKKKKKKKKKK KX

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK X

ABG92685  ck: 3915  len: 57    ! Abg92685 Human DNA-binding protein #111. 11

1          (R,K){20,20}
           (K){20}
31: KQLLL KKKKKKKKKKKKKKKKKKKK KKKXG

           (K){20}
32: QLLLK KKKKKKKKKKKKKKKKKKKK KKXGG

           (K){20}
33: LLLKK KKKKKKKKKKKKKKKKKKKK KXGGF

           (K){20}
34: LLKKK KKKKKKKKKKKKKKKKKKKK XGGF

ABG92686  ck: 3679  len: 37    ! Abg92686 Human DNA-binding protein #112. 11

1          (R,K){20,20}
           (K){20}
15: ISPLT KKKKKKKKKKKKKKKKKKKK KKX

           (K){20}
16: SPLTK KKKKKKKKKKKKKKKKKKKK KX

           (K){20}
17: PLTKK KKKKKKKKKKKKKKKKKKKK X

ABG92687  ck: 657   len: 196   ! Abg92687 Human DNA-binding protein #113. 11

1          (R,K){20,20}
           (K){20}
169: FVXFE KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
170: VXFEK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
171: XFEKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
172: FEKKK KKKKKKKKKKKKKKKKKKKK KKKKX

           (K){20}
173: EKKKK KKKKKKKKKKKKKKKKKKKK KKKX

           (K){20}
174: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

           (K){20}
175: KKKKK KKKKKKKKKKKKKKKKKKKK KX

           (K){20}
176: KKKKK KKKKKKKKKKKKKKKKKKKK X

ABG92688  ck: 4672  len: 57    ! Abg92688 Human DNA-binding protein #114. 11

1          (R,K){20,20}
           (K){20}
28: DKTFH KKKKKKKKKKKKKKKKKKKK KKKKX

           (K){20}
29: KTFHK KKKKKKKKKKKKKKKKKKKK KKKXP

           (K){20}
30: TFHKK KKKKKKKKKKKKKKKKKKKK KKXPG

           (K){20}
31: FHKKK KKKKKKKKKKKKKKKKKKKK KXPGG

           (K){20}
32: HKKKK KKKKKKKKKKKKKKKKKKKK XPGGG

ABG92689  ck: 9656  len: 66    ! Abg92689 Human DNA-binding protein #115. 11

1          (R,K){20,20}
           (K){20}
38: MVISV KKKKKKKKKKKKKKKKKKKK KKKKR

           (K){20}
39: VISVK KKKKKKKKKKKKKKKKKKKK KKKRE

           (K){20}
40: ISVKK KKKKKKKKKKKKKKKKKKKK KKREK

           (K){20}
41: SVKKK KKKKKKKKKKKKKKKKKKKK KREKX

           (K){20}
42: VKKKK KKKKKKKKKKKKKKKKKKKK REKXK

           (R,K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKR EKXK

ABG92691  ck: 4665  len: 34    ! Abg92691 Human DNA-binding protein #117. 11

1          (R,K){20,20}
           (K){20}
10: PELLL KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
11: ELLLK KKKKKKKKKKKKKKKKKKKK KKKK

           (K){20}
12: LLLKK KKKKKKKKKKKKKKKKKKKK KKK

           (K){20}
13: LLKKK KKKKKKKKKKKKKKKKKKKK KK

           (K){20}
14: LKKKK KKKKKKKKKKKKKKKKKKKK K

           (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK

ABG92692  ck: 7810  len: 64    ! Abg92692 Human DNA-binding protein #118. 11

1          (R,K){20,20}
           (K){20}
37: LKYFW KKKKKKKKKKKKKKKKKKKK KKKKX

           (K){20}
38: KYFWK KKKKKKKKKKKKKKKKKKKK KKKXG

           (K){20}
39: YFWKK KKKKKKKKKKKKKKKKKKKK KKXGX

           (K){20}
40: FWKKK KKKKKKKKKKKKKKKKKKKK KXGXP
```

```
            (K){20}
41: WKKKK KKKKKKKKKKKKKKKKKKKK XGXP

ABG92693  ck: 269   len: 76    ! Abg92693 Human DNA-binding protein #119. 11

1          (R,K){20,20}
            (K){20}
37: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
38: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
39: SRAKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
40: RAKKK KKKKKKKKKKKKKKKKKKKK KKKKX

            (K){20}
41: AKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

            (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

            (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KXKKI

            (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK XKKIK

ABG92694  ck: 8370  len: 45    ! Abg92694 Human DNA-binding protein #120. 11

1          (R,K){20,20}
            (K){20}
17: APKTQ KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
18: PKTQK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
19: KTQKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
20: TQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
21: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

            (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKX

            (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

            (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KX

            (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK X

ABG92695  ck: 1663  len: 87    ! Abg92695 Human DNA-binding protein #121. 11

1          (R,K){20,20}
            (K){20}
36: KWSSX KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
37: WSSXK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
38: SSXKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
39: SXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
40: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

            (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

            (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGX

            (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KRGXP

            (K){20}
```

```
62: KKKKK KKKKKKKKKKKKKKKKKKKK RGXPF

          (R,K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKR GXPFX

ABG92696  ck: 5607  len: 63   ! Abg92696 Human DNA-binding protein #122. 11

1         (R,K){20,20}
            (K){20}
26: MVELE KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
27: VELEK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
28: ELEKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
29: LEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
30: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

            (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

            (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

            (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KGGPF

            (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK GGPF

ABG92697  ck: 5997  len: 58   ! Abg92697 Human DNA-binding protein #123. 11

1         (R,K){20,20}
            (K){20}
28: RPTRP KKKKKKKKKKKKKKKKKKKK KXKKK

            (K){20}
29: PTRPK KKKKKKKKKKKKKKKKKKKK XKKKK

ABG92698  ck: 5764  len: 74   ! Abg92698 Human DNA-binding protein #124. 11

1         (R,K){20,20}
            (K){20}
40: EFLSA KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
41: FLSAK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
42: LSAKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
43: SAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
44: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

            (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKX

            (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

            (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KX

            (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK X

ABG92699  ck: 7734  len: 97   ! Abg92699 Human DNA-binding protein #125. 11

1         (R,K){20,20}
            (K){20}
57: RGCSY KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
58: GCSYK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
59: CSYKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
60: SYKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
61: YKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
 67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 70: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 71: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
 72: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXG

                (K){20}
 73: KKKKK KKKKKKKKKKKKKKKKKKKK KKXGX

                (K){20}
 74: KKKKK KKKKKKKKKKKKKKKKKKKK KXGX

                (K){20}
 75: KKKKK KKKKKKKKKKKKKKKKKKKK XGX

 ABG92700   ck: 3164   len: 181    ! Abg92700 Human DNA-binding protein #126. 11

1              (R,K){20,20}
                (K){20}
154: TRKPE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
155: RKPEK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
156: KPEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
157: PEKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
158: EKKKK KKKKKKKKKKKKKKKKKKKK KKKX

                (K){20}
159: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

                (K){20}
160: KKKKK KKKKKKKKKKKKKKKKKKKK KX

                (K){20}
161: KKKKK KKKKKKKKKKKKKKKKKKKK X

 ABG92701   ck: 7117   len: 55     ! Abg92701 Human DNA-binding protein #127. 11

1              (R,K){20,20}
                (K){20}
 22: DDKKN KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 23: DKKNK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
 24: KKNKK KKKKKKKKKKKKKKKKKKKK KKKXG

                (K){20}
 25: KNKKK KKKKKKKKKKKKKKKKKKKK KKXGG

                (K){20}
 26: NKKKK KKKKKKKKKKKKKKKKKKKK KXGGR

                (K){20}
 27: KKKKK KKKKKKKKKKKKKKKKKKKK XGGRF

 ABG92702   ck: 9316   len: 67     ! Abg92702 Human DNA-binding protein #128. 11/

1              (R,K){20,20}
                (R,K){20}
 36: FTLQT RKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 37: TLQTR KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
 38: LQTRK KKKKKKKKKKKKKKKKKKKK KKKXG

                (K){20}
 39: QTRKK KKKKKKKKKKKKKKKKKKKK KKXGG

                (K){20}
 40: TRKKK KKKKKKKKKKKKKKKKKKKK KXGGR

                (K){20}
 41: RKKKK KKKKKKKKKKKKKKKKKKKK XGGRF

 ABG78915   ck: 8085   len: 154    ! Abg78915 Human breast tumour polypeptide #7

1              (R,K){20,20}
                (K){20}
114: TQLRQ KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
115: QLRQK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
116: LRQKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
117: RQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
118: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
121: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
122: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
123: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
124: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
125: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                (K){20}
126: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

                (K){20}
127: KKKKK KKKKKKKKKKKKKKKKKKKK KXKKK
```

```
           (K){20}
128: KKKKK KKKKKKKKKKKKKKKKKKKK XKKKK

 ABJ10296  ck: 904   len: 76    ! Abj10296 Human lung specific protein SEQ ID
1           (R,K){20,20}
             (R,K){20}
  5:  KNKQ KKKRKKRKKRKKRKKRKKRK RKRKK

             (R,K){20}
  6: KNKQK KKRKKRKKRKKRKKRKKRKR KRKKK

             (R,K){20}
  7: NKQKK KRKKRKKRKKRKKRKKRKRK RKKKR

             (R,K){20}
  8: KQKKK RKKRKKRKKRKKRKKRKRKR KKKRR

             (R,K){20}
  9: QKKKR KKRKKRKKRKKRKKRKRKRK KKRRK

             (R,K){20}
 10: KKKRK KRKKRKKRKKRKKRKRKRKK KRRKK

             (R,K){20}
 11: KKRKK RKKRKKRKKRKKRKRKRKKK RRKKG

             (R,K){20}
 12: KRKKR KKRKKRKKRKKRKRKRKKKR RKKGR

             (R,K){20}
 13: RKKRK KRKKRKKRKKRKRKRKKKRR KKGRR

             (R,K){20}
 14: KKRKK RKKRKKRKKRKRKRKKKRRK KGRRR

             (R,K){20}
 15: KRKKR KKRKKRKKRKRKRKKKRRKK GRRRR

             (R,K){20}
 36: RRKKG RRRRKKKKKKKKKKKKKKKK KKKKK

             (R,K){20}
 37: RKKGR RRRKKKKKKKKKKKKKKKKK KKKKK

             (R,K){20}
 38: KKGRR RRKKKKKKKKKKKKKKKKKK KKKKK

             (R,K){20}
 39: KGRRR RKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 40: GRRRR KKKKKKKKKKKKKKKKKKKK KKKKE

              (K){20}
 41: RRRRK KKKKKKKKKKKKKKKKKKKK KKKER

              (K){20}
 42: RRRKK KKKKKKKKKKKKKKKKKKKK KKERK

              (K){20}
 43: RRKKK KKKKKKKKKKKKKKKKKKKK KERKK

              (K){20}
 44: RKKKK KKKKKKKKKKKKKKKKKKKK ERKKE

 ABP62049  ck: 5821  len: 62    ! Abp62049 Human secreted protein SEQ ID NO 1
1           (R,K){20,20}
              (K){20}
 40: QFQAS KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
 41: FQASK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
 42: QASKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
 43: ASKKK KKKKKKKKKKKKKKKKKKKK

 ABG64148  ck: 5584  len: 73    ! Abg64148 Human albumin fusion protein #823.
1           (R,K){20,20}
              (K){20}
 47: LPTFL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 48: PTFLK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 49: TFLKK KKKKKKKKKKKKKKKKKKKK KKKKI

              (K){20}
 50: FLKKK KKKKKKKKKKKKKKKKKKKK KKKI

              (K){20}
 51: LKKKK KKKKKKKKKKKKKKKKKKKK KKI

              (K){20}
 52: KKKKK KKKKKKKKKKKKKKKKKKKK KI

              (K){20}
 53: KKKKK KKKKKKKKKKKKKKKKKKKK I

 ABG64149  ck: 4416  len: 72    ! Abg64149 Human albumin fusion protein #824.
1           (R,K){20,20}
              (K){20}
 47: LPTFL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 48: PTFLK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 49: TFLKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
 50: FLKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
 51: LKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
 52: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
 53: KKKKK KKKKKKKKKKKKKKKKKKKK

 ABG65250  ck: 7046  len: 105   ! Abg65250 Human albumin fusion protein #1925
1           (R,K){20,20}
              (K){20}
 78: TLLVL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 79: LLVLK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 80: LVLKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                 (K){20}
      81: VLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                 (K){20}
      83: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                 (K){20}
      84: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                 (K){20}
      85: KKKKK KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
      86: KKKKK KKKKKKKKKKKKKKKKKKKK

      ABG65251   ck: 8971   len: 108    ! Abg65251 Human albumin fusion protein #1926

1               (R,K){20,20}
                 (K){20}
      78: TLLXL KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      79: LLXLK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      80: LXLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      81: XLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                 (K){20}
      86: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

      ABG65440   ck: 1431   len: 530    ! Abg65440 Human albumin fusion protein #2115

1               (R,K){20,20}
                 (K){20}
     511: LHAPP KKKKKKKKKKKKKKKKKKKK

      ABG35896   ck: 5383   len: 86     ! Abg35896 Human peptide encoded by genome-de

1               (R,K){20,20}
                 (R,K){20}
      15: RRRRG RRRRRKKKKKKRKKKKKRKK KKRRR

                 (R,K){20}
      16: RRRGR RRRRKKKKKKRKKKKKRKKK KRRRR

                 (R,K){20}
      17: RRGRR RRRKKKKKKRKKKKKRKKKK RRRRR

                 (R,K){20}
      18: RGRRR RRKKKKKKRKKKKKRKKKKR RRRRR

                 (R,K){20}
      19: GRRRR RKKKKKKRKKKKKRKKKKRR RRRRR

                 (R,K){20}
      20: RRRRR KKKKKKRKKKKKRKKKKRRR RRRRR

                 (R,K){20}
      21: RRRRK KKKKKRKKKKKRKKKKRRRR RRRRR

                 (R,K){20}
      22: RRRKK KKKKRKKKKKRKKKKRRRRR RRRRR

                 (R,K){20}
      23: RRKKK KKKRKKKKKRKKKKRRRRRR RRRRR

                 (R,K){20}
      24: RKKKK KKRKKKKKRKKKKRRRRRRR RRRRR

                 (R,K){20}
      25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

                 (R,K){20}
      26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

                 (R,K){20}
      27: KKKKR KKKKKRKKKKRRRRRRRRRR RRRRR

                 (R,K){20}
      28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

                 (R,K){20}
      29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

                 (R,K){20}
      30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

                 (R,K){20}
      31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

                 (R,K){20}
      32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

                 (R,K){20}
      33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

                 (R,K){20}
      34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

                 (R,K){20}
      35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

                 (R,K){20}
      36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
      37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
      38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
      39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
      40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
      41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                  (R){20}
      42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

                  (R){20}
      43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ
```

```
            (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

            (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

ABG36760  ck: 1334  len: 86    ! Abg36760 Human peptide encoded by genome-de

1          (R,K){20,20}
            (R,K){20}
57: EEEEG RRRKKKRKRKKKKKKKKKKK KKKKK

            (R,K){20}
58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

            (R,K){20}
59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

            (R,K){20}
60: EGRRR KKKKRKRKKKKKKKKKKKKK KKKKK

            (R,K){20}
61: GRRRK KKKRKRKKKKKKKKKKKKKK KKKKK

            (R,K){20}
62: RRRKK KKRKRKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
63: RRKKK KRKRKKKKKKKKKKKKKKKK KKKK

            (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKKK KKK

            (R,K){20}
65: KKKKR KRKKKKKKKKKKKKKKKKKK KK

            (R,K){20}
66: KKKRK RKKKKKKKKKKKKKKKKKKK K

             (K){20}
67: KKRKR KKKKKKKKKKKKKKKKKKKK

ABG36843  ck: 9082  len: 167   ! Abg36843 Human peptide encoded by genome-de

1          (R,K){20,20}
            (R,K){20}
33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

            (R,K){20}
34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

            (R,K){20}
35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR

ABG37280  ck: 2276  len: 89    ! Abg37280 Human peptide encoded by genome-de

1          (R,K){20,20}
            (R,K){20}
23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

             (K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

             (K){20}
47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

            (R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

            (R,K){20}
49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

ABG37848  ck: 3607  len: 88    ! Abg37848 Human peptide encoded by genome-de

1          (R,K){20,20}
             (K){20}
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK

ABG38450  ck: 3937  len: 85    ! Abg38450 Human peptide encoded by genome-de

1           (R,K){20,20}
              (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

              (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

              (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

              (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

              (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

ABG39191  ck: 2686  len: 71    ! Abg39191 Human peptide encoded by genome-de

1           (R,K){20,20}
              (K){20}
20: KKKKQ KKKKKKKKKKKKKKKKKKKK KKKKR

              (K){20}
21: KKKQK KKKKKKKKKKKKKKKKKKKK KKKRK

              (K){20}
22: KKQKK KKKKKKKKKKKKKKKKKKKK KKRKR

              (K){20}
23: KQKKK KKKKKKKKKKKKKKKKKKKK KRKRK

              (K){20}
24: QKKKK KKKKKKKKKKKKKKKKKKKK RKRKK

            (R,K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKR KRKKK

            (R,K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKRK RKKKK

            (R,K){20}
27: KKKKK KKKKKKKKKKKKKKKKKRKR KKKKK

            (R,K){20}
28: KKKKK KKKKKKKKKKKKKKKKRKRK KKKKK

            (R,K){20}
29: KKKKK KKKKKKKKKKKKKKKRKRKK KKKKK

            (R,K){20}
30: KKKKK KKKKKKKKKKKKKKRKRKKK KKKKK

            (R,K){20}
31: KKKKK KKKKKKKKKKKKKRKRKKKK KKKKK

            (R,K){20}
32: KKKKK KKKKKKKKKKKKRKRKKKKK KKKKK

            (R,K){20}
33: KKKKK KKKKKKKKKKKRKRKKKKKK KKKKK

            (R,K){20}
34: KKKKK KKKKKKKKKKRKRKKKKKKK KKKKK

            (R,K){20}
35: KKKKK KKKKKKKKKRKRKKKKKKKK KKKKK

            (R,K){20}
36: KKKKK KKKKKKKKRKRKKKKKKKKK KKKKK

            (R,K){20}
37: KKKKK KKKKKKKRKRKKKKKKKKKK KKKKK

            (R,K){20}
38: KKKKK KKKKKKRKRKKKKKKKKKKK KKKKK

            (R,K){20}
39: KKKKK KKKKKRKRKKKKKKKKKKKK KKKKK

            (R,K){20}
40: KKKKK KKKKRKRKKKKKKKKKKKKK KKKKK

            (R,K){20}
41: KKKKK KKKRKRKKKKKKKKKKKKKK KKKKK

            (R,K){20}
42: KKKKK KKRKRKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
43: KKKKK KRKRKKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
44: KKKKK RKRKKKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
45: KKKKR KRKKKKKKKKKKKKKKKKKK KKKKS
```

```
          (R,K){20}
46: KKKRK RKKKKKKKKKKKKKKKKKKK KKKSA

           (K){20}
47: KKRKR KKKKKKKKKKKKKKKKKKKK KKSAH

           (K){20}
48: KRKRK KKKKKKKKKKKKKKKKKKKK KSAH

           (K){20}
49: RKRKK KKKKKKKKKKKKKKKKKKKK SAH

ABG40383  ck: 1560  len: 88    ! Abg40383 Human peptide encoded by genome-de

1         (R,K){20,20}
           (R){20}
43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

           (R){20}
59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

           (R){20}
60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

           (R){20}
61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

           (R){20}
62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

ABG40676  ck: 2324  len: 36    ! Abg40676 Human peptide encoded by genome-de

1         (R,K){20,20}
          (R,K){20}
16: KERKT KKRKRKRRRRRKKRRRKKRR R

          (R,K){20}
17: ERKTK KRKRKRRRRRKKRRRKKRRR

ABG43651  ck: 8343  len: 66    ! Abg43651 Human peptide encoded by genome-de

1         (R,K){20,20}
          (R,K){20}
 6: ETERE KKKKRKKKKKKKKKKKKKKK KKKNK

          (R,K){20}
 7: TEREK KKKRKKKKKKKKKKKKKKKK KKNKK

          (R,K){20}
 8: EREKK KKRKKKKKKKKKKKKKKKKK KNKKK

          (R,K){20}
 9: REKKK KRKKKKKKKKKKKKKKKKKK NKKKK

ABG45640  ck: 2394  len: 57    ! Abg45640 Human peptide encoded by genome-de

1         (R,K){20,20}
           (R){20}
20: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
21: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRG

           (R){20}
22: EEGRR RRRRRRRRRRRRRRRRRRRR RRRGR

           (R){20}
23: EGRRR RRRRRRRRRRRRRRRRRRRR RRGRR

           (R){20}
24: GRRRR RRRRRRRRRRRRRRRRRRRR RGRRR

           (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR

ABG46635  ck: 3301  len: 52    ! Abg46635 Human peptide encoded by genome-de

1         (R,K){20,20}
          (R,K){20}
12: KKKKN KKKKKKKKKRKRKRKRKKKK KKKKK

          (R,K){20}
13: KKKNK KKKKKKKKRKRKRKRKKKKK KKKKK

          (R,K){20}
14: KKNKK KKKKKKKRKRKRKRKKKKKK KKKKK

          (R,K){20}
15: KNKKK KKKKKKRKRKRKRKKKKKKK KKKKK

          (R,K){20}
16: NKKKK KKKKKRKRKRKRKKKKKKKK KKKKK

          (R,K){20}
17: KKKKK KKKKRKRKRKRKKKKKKKKK KKKKK
```

```
              (R,K){20}
    18: KKKKK KKKRKRKRKRKKKKKKKKKK KKKKK

              (R,K){20}
    19: KKKKK KKRKRKRKRKKKKKKKKKKK KKKKK

              (R,K){20}
    20: KKKKK KRKRKRKRKKKKKKKKKKKK KKKKK

              (R,K){20}
    21: KKKKK RKRKRKRKKKKKKKKKKKKK KKKKK

              (R,K){20}
    22: KKKKR KRKRKRKKKKKKKKKKKKKK KKKKK

              (R,K){20}
    23: KKKRK RKRKRKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
    24: KKRKR KRKRKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
    25: KRKRK RKRKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
    26: RKRKR KRKKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
    27: KRKRK RKKKKKKKKKKKKKKKKKKK KKKKA

               (K){20}
    28: RKRKR KKKKKKKKKKKKKKKKKKKK KKKAF

               (K){20}
    29: KRKRK KKKKKKKKKKKKKKKKKKKK KKAF

               (K){20}
    30: RKRKK KKKKKKKKKKKKKKKKKKKK KAF

               (K){20}
    31: KRKKK KKKKKKKKKKKKKKKKKKKK AF

    ABG47062  ck: 4895  len: 51    ! Abg47062 Human peptide encoded by genome-de

1            (R,K){20,20}
              (R,K){20}
    18: LFKPM RKRRRKKKRRRKKRKKKRRR KKLTT

              (R,K){20}
    19: FKPMR KRRRKKKRRRKKRKKKRRRK KLTTT

              (R,K){20}
    20: KPMRK RRRKKKRRRKKRKKKRRRKK LTTTT

    ABB77424  ck: 9633  len: 637   ! Abb77424 Human tumour marker protein se2-5.

1            (R,K){20,20}
              (R,K){20}
   618: GDKTD RKKKKKKKKKKKKKKKKKKK

    AAE20631  ck: 8971  len: 108   ! Aae20631 Human gene 4 encoded secreted prot

1            (R,K){20,20}
               (K){20}
    78: TLLXL KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    79: LLXLK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    80: LXLKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    81: XLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

               (K){20}
    83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

               (K){20}
    84: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

               (K){20}
    85: KKKKK KKKKKKKKKKKKKKKKKKKK KXKK

               (K){20}
    86: KKKKK KKKKKKKKKKKKKKKKKKKK XKK

    AAE20642  ck: 7046  len: 105   ! Aae20642 Human gene 4 encoded secreted prote

1            (R,K){20,20}
               (K){20}
    78: TLLVL KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    79: LLVLK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    80: LVLKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    81: VLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKK

               (K){20}
    83: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
    84: KKKKK KKKKKKKKKKKKKKKKKKKK KK

               (K){20}
    85: KKKKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
    86: KKKKK KKKKKKKKKKKKKKKKKKKK

    ABB89690  ck: 1158  len: 226   ! Abb89690 Human polypeptide SEQ ID NO 2066. S

1            (R,K){20,20}
               (K){20}
   198: GAESL KKKKKKKKKKKKKKKKKKKK KKGRP

               (K){20}
   199: AESLK KKKKKKKKKKKKKKKKKKKK KGRPX

               (K){20}
   200: ESLKK KKKKKKKKKKKKKKKKKKKK GRPXX

    AAE14544  ck: 7250  len: 35    ! Aae14544 Peptide p65 used in assay for detec

1            (R,K){20,20}
               (K){20}
     1:       KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
     2:     K KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
           (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

           (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

           (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KK

           (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK K

           (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK

AAU75162  ck: 5536  len: 84    ! Aau75162 Single-chain antigen-binding polyp

1         (R,K){20,20}
           (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX
```

```
              (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXX

              (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKXXX

              (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KXXXX

              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK XXXXX

AAU75163  ck: 8024  len: 84    ! Aau75163 Single-chain antigen-binding polyp

1            (R,K){20,20}
              (R){20}
 1:       RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 2:     R RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 3:    RR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 4:   RRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 5:  RRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 6: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 7: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 8: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 9: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
10: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
11: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
12: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
13: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
14: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
15: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
16: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
17: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
18: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
19: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
20: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
21: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
22: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
23: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
24: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
26: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
27: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
28: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
29: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
30: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
31: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
32: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
33: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRX

              (R){20}
34: RRRRR RRRRRRRRRRRRRRRRRRRR RRRXX

              (R){20}
35: RRRRR RRRRRRRRRRRRRRRRRRRR RRXXX

              (R){20}
36: RRRRR RRRRRRRRRRRRRRRRRRRR RXXXX

              (R){20}
37: RRRRR RRRRRRRRRRRRRRRRRRRR XXXXX

AAU75164  ck: 9643  len: 83    ! Aau75164 Single-chain antigen-binding polyp

1            (R,K){20,20}
              (R,K){20}
 1:       RKRKRKRKRKRKRKRKRKRK RKRKR

              (R,K){20}
 2:     R KRKRKRKRKRKRKRKRKRKR KRKRK

              (R,K){20}
 3:    RK RKRKRKRKRKRKRKRKRKRK RKRKR

              (R,K){20}
 4:   RKR KRKRKRKRKRKRKRKRKRKR KRKRK

              (R,K){20}
 5:  RKRK RKRKRKRKRKRKRKRKRKRK RKRKR
```

```
               (R,K){20}
 6: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
 7: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
 8: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
 9: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
10: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
11: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
12: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
13: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
14: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
15: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
16: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
17: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
18: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
19: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
20: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
21: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
22: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
23: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
24: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
25: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
26: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
27: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
28: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
29: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
30: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
31: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

               (R,K){20}
32: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

               (R,K){20}
33: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKX

               (R,K){20}
34: RKRKR KRKRKRKRKRKRKRKRKRKR KRKXX

               (R,K){20}
35: KRKRK RKRKRKRKRKRKRKRKRKRK RKXXX

               (R,K){20}
36: RKRKR KRKRKRKRKRKRKRKRKRKR KXXXX

               (R,K){20}
37: KRKRK RKRKRKRKRKRKRKRKRKRK XXXXX

AAU75165   ck: 6708   len: 84    ! Aau75165 Single-chain antigen-binding polyp

1              (R,K){20,20}
                (R){20}
 1:       RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
 2:     R RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
 3:    RR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
 4:   RRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
 5:  RRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
 6: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
 7: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
 8: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
 9: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
10: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
11: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
12: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
13: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
14: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
15: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
```

```
16: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
17: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
18: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
19: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
20: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
21: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
22: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
23: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
24: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
26: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
27: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
28: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
29: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
30: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
31: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
32: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
33: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRX

          (R){20}
34: RRRRR RRRRRRRRRRRRRRRRRRRR RRRXX

          (R){20}
35: RRRRR RRRRRRRRRRRRRRRRRRRR RRXXX

          (R){20}
36: RRRRR RRRRRRRRRRRRRRRRRRRR RXXXX

          (R){20}
37: RRRRR RRRRRRRRRRRRRRRRRRRR XXXXX

AAE16634   ck: 1504  len: 32    ! Aae16634 Nuclear localisation signal peptid

1               (R,K){20,20}
                 (R,K){20}
 1:             KRKRKRKRKRKRKRKRKRKR KRKRK

                 (R,K){20}

 2:     K RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
 3:    KR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
 4:   KRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
 5:  KRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
 6: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
 7: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
 8: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
 9: RKRKR KRKRKRKRKRKRKRKRKRKR KRKR

          (R,K){20}
10: KRKRK RKRKRKRKRKRKRKRKRKRK RKR

          (R,K){20}
11: RKRKR KRKRKRKRKRKRKRKRKRKR KR

          (R,K){20}
12: KRKRK RKRKRKRKRKRKRKRKRKRK R

          (R,K){20}
13: RKRKR KRKRKRKRKRKRKRKRKRKR

AAM48798   ck: 6221  len: 26    ! Aam48798 Tumour-targetting peptide vector pe

1        (R,K){20,20}
           (K){20}
 7: RGDCF KKKKKKKKKKKKKKKKKKKK

ABB44830   ck: 5275  len: 38    ! Abb44830 Human protective sequence CNI-0074

1        (R,K){20,20}
           (K){20}
 4:   MAQ KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 5:  MAQK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 6: MAQKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: AQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 8: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
            (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

            (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

            (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKGG

            (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KGG

            (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK GG

AAU69690  ck: 875  len: 46   ! Aau69690 Cell death protective sequence CNI

1          (R,K){20,20}
            (K){20}
15: REKKS KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
16: EKKSK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
17: KKSKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
18: KSKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
19: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

            (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

            (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKGG

            (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KGG

            (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK GG

AAU69736  ck: 5862  len: 50   ! Aau69736 Cell death protective sequence CNI

1          (R,K){20,20}
            (R,K){20}
11: ESALG RKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
12: SALGR KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
13: ALGRK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
14: LGRKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
15: GRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
16: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

            (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGR

            (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKGR

            (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KGR

            (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK GR

ABG47914  ck: 5383  len: 86   ! Abg47914 Human liver peptide, SEQ ID No 265

1          (R,K){20,20}
            (R,K){20}
15: RRRRG RRRRRKKKKKRKKKKKRKK KKRRR

            (R,K){20}
16: RRRGR RRRRKKKKKRKKKKKRKKK KRRRR

            (R,K){20}
17: RRGRR RRRKKKKKRKKKKKRKKKK RRRRR

            (R,K){20}
18: RGRRR RRKKKKKRKKKKKRKKKKR RRRRR

            (R,K){20}
19: GRRRR RKKKKKRKKKKKRKKKKRR RRRRR

            (R,K){20}
20: RRRRR KKKKKRKKKKKRKKKKRRR RRRRR

            (R,K){20}
21: RRRRK KKKKRKKKKKRKKKKRRRR RRRRR

            (R,K){20}
22: RRRKK KKKRKKKKKRKKKKRRRRR RRRRR

            (R,K){20}
23: RRKKK KKRKKKKKRKKKKRRRRRR RRRRR

            (R,K){20}
24: RKKKK KRKKKKKRKKKKRRRRRRR RRRRR
```

```
              (R,K){20}
25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

              (R,K){20}
26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

              (R,K){20}
27: KKKKR KKKKKRKKKKRRRRRRRRRR RRRRR

              (R,K){20}
28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

              (R,K){20}
29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

              (R,K){20}
30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

              (R,K){20}
31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

              (R,K){20}
32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

              (R,K){20}
33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

              (R,K){20}
34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

              (R,K){20}
35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

              (R,K){20}
36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

               (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

               (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

               (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

               (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

ABG48766  ck: 1334  len: 86   ! Abg48766 Human liver peptide, SEQ ID No 274

1            (R,K){20,20}
              (R,K){20}
57: EEEEG RRRKKKKRKRKKKKKKKKKK KKKKK

              (R,K){20}
58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

              (R,K){20}
59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

              (R,K){20}
60: EGRRR KKKKRKRKKKKKKKKKKKKK KKKKK

              (R,K){20}
61: GRRRK KKKRKRKKKKKKKKKKKKKK KKKKK

              (R,K){20}
62: RRRKK KKRKRKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
63: RRKKK KRKRKKKKKKKKKKKKKKKK KKKK

              (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKKK KKK

              (R,K){20}
65: KKKKR KRKKKKKKKKKKKKKKKKKK KK

              (R,K){20}
66: KKKRK RKKKKKKKKKKKKKKKKKKK K

               (K){20}
67: KKRKR KKKKKKKKKKKKKKKKKKKK

ABG48847  ck: 9082  len: 167   ! Abg48847 Human liver peptide, SEQ ID No 274

1            (R,K){20,20}
              (R,K){20}
33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

              (R,K){20}
34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

              (R,K){20}
35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR

ABG49354  ck: 2276  len: 89   ! Abg49354 Human liver peptide, SEQ ID No 280

1            (R,K){20,20}
              (R,K){20}
23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

               (K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

               (K){20}
47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

              (R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

              (R,K){20}
49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

ABG49651  ck: 1939  len: 130   ! Abg49651 Human liver peptide, SEQ ID No 282

1            (R,K){20,20}
              (R,K){20}
42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR

              (R,K){20}
43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

              (R,K){20}
44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK
```

```
               (R,K){20}
45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

               (R,K){20}
46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

               (R,K){20}
47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

               (R,K){20}
48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

               (R,K){20}
49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

               (R,K){20}
50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

               (R,K){20}
51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

               (R,K){20}
52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

               (R,K){20}
53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE

               (R,K){20}
54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE

ABG49962  ck: 3607  len: 88    ! Abg49962 Human liver peptide, SEQ ID No 286

1             (R,K){20,20}
                (K){20}
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK

ABG50530  ck: 3937  len: 85    ! Abg50530 Human liver peptide, SEQ ID No 291

1             (R,K){20,20}
                (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                 (K){20}
  6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
  7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

                 (K){20}
  8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

                 (K){20}
  9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

                 (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

                 (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

ABG52382  ck: 1560  len: 88   ! Abg52382 Human liver peptide, SEQ ID No 310

1               (R,K){20,20}
                 (R){20}
 43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                 (R){20}
 59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

                 (R){20}
 60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

                 (R){20}
 61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

                 (R){20}
 62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

ABG52597  ck: 2324  len: 36   ! Abg52597 Human liver peptide, SEQ ID No 312

1               (R,K){20,20}
                 (R,K){20}
 16: KERKT KKRKRKRRRRRKKRRRKKRR R

                 (R,K){20}
 17: ERKTK KRKRKRRRRRKKRRRKKRRR

ABG55512  ck: 8343  len: 66   ! Abg55512 Human liver peptide, SEQ ID No 341

1               (R,K){20,20}
                 (R,K){20}
  6: ETERE KKKKRKKKKKKKKKKKKKKK KKKNK

                 (R,K){20}
  7: TEREK KKKRKKKKKKKKKKKKKKKK KKNKK

                 (R,K){20}
  8: EREKK KKRKKKKKKKKKKKKKKKKK KNKKK

                 (R,K){20}
  9: REKKK KRKKKKKKKKKKKKKKKKKK NKKKK

ABG58055  ck: 2394  len: 57   ! Abg58055 Human liver peptide, SEQ ID No 367

1               (R,K){20,20}
                 (R){20}
 20: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 21: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRG

                 (R){20}
 22: EEGRR RRRRRRRRRRRRRRRRRRRR RRRGR

                 (R){20}
 23: EGRRR RRRRRRRRRRRRRRRRRRRR RRGRR

                 (R){20}
 24: GRRRR RRRRRRRRRRRRRRRRRRRR RGRRR

                 (R){20}
 25: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR

ABG58577  ck: 4228  len: 24   ! Abg58577 Human liver peptide, SEQ ID No 372

1               (R,K){20,20}
                 (R,K){20}
  1:       RRRRRRRKKKKRRRRRRRRR RKRT

                 (R,K){20}
  2:     R RRRRRRKKKKRRRRRRRRRR KRT

                 (R,K){20}
  3:    RR RRRRRKKKKRRRRRRRRRRK RT

                 (R,K){20}
  4:   RRR RRRRKKKKRRRRRRRRRRKR T
```

```
          ABG59254  ck: 3301  len: 52    ! Abg59254 Human liver peptide, SEQ ID No 379
1                   (R,K){20,20}
                     (R,K){20}
          12: KKKKN KKKKKKKKKRKRKRKRKKKK KKKKK

                     (R,K){20}
          13: KKKNK KKKKKKKKRKRKRKRKKKKK KKKKK

                     (R,K){20}
          14: KKNKK KKKKKKKRKRKRKRKKKKKK KKKKK

                     (R,K){20}
          15: KNKKK KKKKKKRKRKRKRKKKKKKK KKKKK

                     (R,K){20}
          16: NKKKK KKKKKRKRKRKRKKKKKKKK KKKKK

                     (R,K){20}
          17: KKKKK KKKKRKRKRKRKKKKKKKKK KKKKK

                     (R,K){20}
          18: KKKKK KKKRKRKRKRKKKKKKKKKK KKKKK

                     (R,K){20}
          19: KKKKK KKRKRKRKRKKKKKKKKKKK KKKKK

                     (R,K){20}
          20: KKKKK KRKRKRKRKKKKKKKKKKKK KKKKK

                     (R,K){20}
          21: KKKKK RKRKRKRKKKKKKKKKKKKK KKKKK

                     (R,K){20}
          22: KKKKR KRKRKRKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
          23: KKKRK RKRKRKKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
          24: KKRKR KRKRKKKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
          25: KRKRK RKRKKKKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
          26: RKRKR KRKKKKKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
          27: KRKRK RKKKKKKKKKKKKKKKKKKK KKKKA

                      (K){20}
          28: RKRKR KKKKKKKKKKKKKKKKKKKK KKKAF

                      (K){20}
          29: KRKRK KKKKKKKKKKKKKKKKKKKK KKAF

                      (K){20}
          30: RKRKK KKKKKKKKKKKKKKKKKKKK KAF

                      (K){20}
          31: KRKKK KKKKKKKKKKKKKKKKKKKK AF

          ABG59685  ck: 4895  len: 51    ! Abg59685 Human liver peptide, SEQ ID No 383
1                   (R,K){20,20}
                     (R,K){20}
          18: LFKPM RKRRRKKKRRRKKRKKKRRR KKLTT

                     (R,K){20}
          19: FKPMR KRRRKKKRRRKKRKKKRRRK KLTTT

                     (R,K){20}
          20: KPMRK RRRKKKRRRKKRKKKRRRKK LTTTT

          ABB96057  ck: 5626  len: 139   ! Abb96057 Human testicular antigen SEQ ID NC
1                   (R,K){20,20}
                      (K){20}
         111: IHLNL KKKKKKKKKKKKKKKKKKKK KKKXK

                      (K){20}
         112: HLNLK KKKKKKKKKKKKKKKKKKKK KKXKK

                      (K){20}
         113: LNLKK KKKKKKKKKKKKKKKKKKKK KXKKK

                      (K){20}
         114: NLKKK KKKKKKKKKKKKKKKKKKKK XKKKK

          ABB96575  ck: 4751  len: 80    ! Abb96575 Human testicular antigen SEQ ID NC
1                   (R,K){20,20}
                      (K){20}
          61: KKXFD KKKKKKKKKKKKKKKKKKKK

          AAU87124  ck: 5599  len: 281   ! Aau87124 Novel central nervous system prote
1                   (R,K){20,20}
                      (K){20}
         248: KKKKQ KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         249: KKKQK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         250: KKQKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         251: KQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         252: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         253: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         254: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         255: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         256: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         257: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
         258: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                      (K){20}
         259: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                      (K){20}
         260: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                      (K){20}
         261: KKKKK KKKKKKKKKKKKKKKKKKKK K
```

```
                 (K){20}
        262: KKKKK KKKKKKKKKKKKKKKKKKKK

         ABG00401  ck: 5116  len: 1,074 ! Abg00401 Novel human diagnostic protein #39
1                (R,K){20,20}
                  (R,K){20}
        609: RGSSS KKKKRRKKRKKKKRKKRKRR KNRKK

                  (R,K){20}
        610: GSSSK KKKRRKKRKKKKRKKRKRRK NRKKK

         ABG03974  ck: 5341  len: 99    ! Abg03974 Novel human diagnostic protein #39
1                (R,K){20,20}
                  (K){20}
          2:     M KKKKKKKKKKKKKKKKKKKK KKKKN

                  (K){20}
          3:    MK KKKKKKKKKKKKKKKKKKKK KKKNS

                  (K){20}
          4:   MKK KKKKKKKKKKKKKKKKKKKK KKNSQ

                  (K){20}
          5:  MKKK KKKKKKKKKKKKKKKKKKKK KNSQI

                  (K){20}
          6: MKKKK KKKKKKKKKKKKKKKKKKKK NSQID

         ABG04391  ck: 9047  len: 139   ! Abg04391 Novel human diagnostic protein #43
1                (R,K){20,20}
                  (K){20}
         79: EEEEE KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         80: EEEEK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         81: EEEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         82: EEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         83: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         84: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         85: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         86: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         87: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         88: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         89: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         90: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}

         91: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         92: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        105: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        106: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        107: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        108: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        109: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

                  (K){20}
        110: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEN

                  (K){20}
        111: KKKKK KKKKKKKKKKKKKKKKKKKK KKENP

                  (K){20}
        112: KKKKK KKKKKKKKKKKKKKKKKKKK KENPF

                  (K){20}
        113: KKKKK KKKKKKKKKKKKKKKKKKKK ENPFS

         ABG05352  ck: 1276  len: 204   ! Abg05352 Novel human diagnostic protein #53
1                (R,K){20,20}
                  (K){20}
```

```
106: EEEEE KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
107: EEEEK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
108: EEEKK KKKKKKKKKKKKKKKKKKKK KKKKR

           (K){20}
109: EEKKK KKKKKKKKKKKKKKKKKKKK KKKRK

           (K){20}
110: EKKKK KKKKKKKKKKKKKKKKKKKK KKRKK

           (K){20}
111: KKKKK KKKKKKKKKKKKKKKKKKKK KRKKK

           (K){20}
112: KKKKK KKKKKKKKKKKKKKKKKKKK RKKKK

          (R,K){20}
113: KKKKK KKKKKKKKKKKKKKKKKKKR KKKKK

          (R,K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKRK KKKKE

          (R,K){20}
115: KKKKK KKKKKKKKKKKKKKKKKRKK KKKEE

          (R,K){20}
116: KKKKK KKKKKKKKKKKKKKKKRKKK KKEEE

          (R,K){20}
117: KKKKK KKKKKKKKKKKKKKKRKKKK KEEEE

          (R,K){20}
118: KKKKK KKKKKKKKKKKKKKRKKKKK EEEEK

 ABG05367   ck: 6907   len: 808    ! Abg05367 Novel human diagnostic protein #53

1         (R,K){20,20}
          (R,K){20}
219: EEEEE RRKKKKKKKKKRRRRRRRRRR RRRRR

          (R,K){20}
220: EEEER RKKKKKKKKRRRRRRRRRRR RRRRR

          (R,K){20}
221: EEERR KKKKKKKKRRRRRRRRRRRR RRRRK

          (R,K){20}
222: EERRK KKKKKKKRRRRRRRRRRRRR RRRKK

          (R,K){20}
223: ERRKK KKKKKKRRRRRRRRRRRRRR RRKKK

          (R,K){20}
224: RRKKK KKKKKRRRRRRRRRRRRRRR RKKKK

          (R,K){20}
225: RKKKK KKKKRRRRRRRRRRRRRRRR KKKKK

          (R,K){20}
226: KKKKK KKKRRRRRRRRRRRRRRRRK KKKKK

          (R,K){20}
227: KKKKK KKRRRRRRRRRRRRRRRRKK KKKKK

          (R,K){20}
228: KKKKK KRRRRRRRRRRRRRRRRKKK KKKKK

          (R,K){20}
229: KKKKK RRRRRRRRRRRRRRRRKKKK KKKKK

          (R,K){20}
230: KKKKR RRRRRRRRRRRRRRRKKKKK KKKKK

          (R,K){20}
231: KKKRR RRRRRRRRRRRRRRKKKKKK KKKKK

          (R,K){20}
232: KKRRR RRRRRRRRRRRRRKKKKKKK KKKKK

          (R,K){20}
233: KRRRR RRRRRRRRRRRRKKKKKKKK KKKKN

          (R,K){20}
234: RRRRR RRRRRRRRRRRKKKKKKKKK KKKNK

          (R,K){20}
235: RRRRR RRRRRRRRRRKKKKKKKKKK KKNKK

          (R,K){20}
236: RRRRR RRRRRRRRRKKKKKKKKKKK KNKKK

          (R,K){20}
237: RRRRR RRRRRRRRKKKKKKKKKKKK NKKKK

           (K){20}
258: KKKKN KKKKKKKKKKKKKKKKKKKK KKNKK

           (K){20}
259: KKKNK KKKKKKKKKKKKKKKKKKKK KNKKN

           (K){20}
260: KKNKK KKKKKKKKKKKKKKKKKKKK NKKNK

 ABG06375   ck: 7807   len: 2,570 ! Abg06375 Novel human diagnostic protein #63

1         (R,K){20,20}
          (R,K){20}
205: KKKSS RRRRRRRKKKRKKKKKKKRR RRRQQ

          (R,K){20}
206: KKSSR RRRRRRKKKRKKKKKKKRRR RRQQQ

          (R,K){20}
207: KSSRR RRRRRKKKRKKKKKKKRRRR RQQQQ

          (R,K){20}
208: SSRRR RRRRKKKRKKKKKKKRRRRR QQQQQ

 ABG06513   ck: 2934   len: 154    ! Abg06513 Novel human diagnostic protein #65

1         (R,K){20,20}
          (R,K){20}
 49: KKKKT RKKRKKRKKKRKKKKKKKKK KKKKK

          (R,K){20}
 50: KKKTR KKRKKRKKKRKKKKKKKKKK KKKKK

          (R,K){20}
 51: KKTRK KRKKRKKKRKKKKKKKKKKK KKKKK

          (R,K){20}
 52: KTRKK RKKRKKKRKKKKKKKKKKKK KKKKK

          (R,K){20}
 53: TRKKR KKRKKKRKKKKKKKKKKKKK KKKKK

          (R,K){20}
 54: RKKRK KRKKKRKKKKKKKKKKKKKK KKKKK
```

```
              (R,K){20}
55: KKRKK RKKKRKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
56: KRKKR KKKRKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
57: RKKRK KKRKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
58: KKRKK KRKKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
59: KRKKK RKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
60: RKKKR KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
61: KKKRK KKKKKKKKKKKKKKKKKKKK KKKKS

               (K){20}
62: KKRKK KKKKKKKKKKKKKKKKKKKK KKKSE

               (K){20}
63: KRKKK KKKKKKKKKKKKKKKKKKKK KKSEX

               (K){20}
64: RKKKK KKKKKKKKKKKKKKKKKKKK KSEXD

               (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK SEXDR

ABG07742  ck: 8672  len: 502   ! Abg07742 Novel human diagnostic protein #77

1            (R,K){20,20}
              (R,K){20}
54: DDEEE RRRKKKKKKKKKKKKKRKKK KKKRK

              (R,K){20}
55: DEEER RRKKKKKKKKKKKKKRKKKK KKRKK

              (R,K){20}
56: EEERR RKKKKKKKKKKKKKRKKKKK KRKKK

              (R,K){20}
57: EERRR KKKKKKKKKKKKKRKKKKKK RKKKK

              (R,K){20}
58: ERRRK KKKKKKKKKKKKRKKKKKKR KKKKK

              (R,K){20}
59: RRRKK KKKKKKKKKKKRKKKKKKRK KKKKK

              (R,K){20}
60: RRKKK KKKKKKKKKKRKKKKKKRKK KKKKK

              (R,K){20}
61: RKKKK KKKKKKKKKRKKKKKKRKKK KKKKK

              (R,K){20}
62: KKKKK KKKKKKKKRKKKKKKRKKKK KKKKK

              (R,K){20}
63: KKKKK KKKKKKKRKKKKKKRKKKKK KKKKN

              (R,K){20}
64: KKKKK KKKKKKRKKKKKKRKKKKKK KKKNI

              (R,K){20}
65: KKKKK KKKKKRKKKKKKRKKKKKKK KKNIL

              (R,K){20}
66: KKKKK KKKKRKKKKKKRKKKKKKKK KNILY

              (R,K){20}
67: KKKKK KKKRKKKKKKRKKKKKKKKK NILYR

ABG10052  ck: 7107  len: 39   ! Abg10052 Novel human diagnostic protein #10

1            (R,K){20,20}
               (K){20}
 2:     Q KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 3:    QK KKKKKKKKKKKKKKKKKKKK KKKKE

               (K){20}
 4:   QKK KKKKKKKKKKKKKKKKKKKK KKKEK

               (K){20}
 5:  QKKK KKKKKKKKKKKKKKKKKKKK KKEKK

               (K){20}
 6: QKKKK KKKKKKKKKKKKKKKKKKKK KEKKK

               (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK EKKKK

ABG10053  ck: 3274  len: 189   ! Abg10053 Novel human diagnostic protein #10

1            (R,K){20,20}
              (R,K){20}
70: EKEKE KKRKKKRKKKKKKKKKKKKK KKKKK

              (R,K){20}
71: KEKEK KRKKKRKKKKKKKKKKKKKK KKKKK

              (R,K){20}
72: EKEKK RKKKRKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
73: KEKKR KKKRKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
74: EKKRK KKRKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
75: KKRKK KRKKKKKKKKKKKKKKKKKK KKKKE

              (R,K){20}
76: KRKKK RKKKKKKKKKKKKKKKKKKK KKKEE

               (K){20}
77: RKKKR KKKKKKKKKKKKKKKKKKKK KKEEE

               (K){20}
78: KKKRK KKKKKKKKKKKKKKKKKKKK KEEEE

               (K){20}
79: KKRKK KKKKKKKKKKKKKKKKKKKK EEEEE

ABG11241  ck: 3870  len: 121   ! Abg11241 Novel human diagnostic protein #11

1            (R,K){20,20}
               (K){20}
44: KKKEE KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
45: KKEEK KKKKKKKKKKKKKKKKKKKK KKKKX

               (K){20}
```

```
46: KEEKK KKKKKKKKKKKKKKKKKKKK KKKXE

            (K){20}
47: EEKKK KKKKKKKKKKKKKKKKKKKK KKXEK

            (K){20}
48: EKKKK KKKKKKKKKKKKKKKKKKKK KXEKE

            (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK XEKEK

ABG11242   ck: 4493   len: 100    ! Abg11242 Novel human diagnostic protein #11

1          (R,K){20,20}
            (K){20}
42: ETPSE KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
43: TPSEK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
44: PSEKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
45: SEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
46: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKN

            (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKNP

            (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK KKNPI

            (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKK KNPIF

            (K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKK NPIFF

ABG11245   ck: 2517   len: 85    ! Abg11245 Novel human diagnostic protein #11

1          (R,K){20,20}
            (K){20}
33: EEEEE KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
34: EEEEK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
35: EEEKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
36: EEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
37: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
```

```
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

            (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGR

            (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKGRK

            (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KGRKT

            (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK GRKTD

ABG11250  ck: 4343  len: 92    ! Abg11250 Novel human diagnostic protein #11

1          (R,K){20,20}
            (K){20}
53: EKEKE KKKKKKKKKKKKKKKKKKKK KKKRK

            (K){20}
54: KEKEK KKKKKKKKKKKKKKKKKKKK KKRKK

            (K){20}
55: EKEKK KKKKKKKKKKKKKKKKKKKK KRKKE

            (K){20}
56: KEKKK KKKKKKKKKKKKKKKKKKKK RKKEM

           (R,K){20}
57: EKKKK KKKKKKKKKKKKKKKKKKKR KKEMA

           (R,K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKRK KEMAL

           (R,K){20}
59: KKKKK KKKKKKKKKKKKKKKKKRKK EMALQ

ABG11266  ck: 9563  len: 146   ! Abg11266 Novel human diagnostic protein #11

1          (R,K){20,20}
            (K){20}
36: KEKRE KKKKKKKKKKKKKKKKKKKK KKKKE

            (K){20}
37: EKREK KKKKKKKKKKKKKKKKKKKK KKKEE

            (K){20}
38: KREKK KKKKKKKKKKKKKKKKKKKK KKEEE

            (K){20}
39: REKKK KKKKKKKKKKKKKKKKKKKK KEEEE

            (K){20}
40: EKKKK KKKKKKKKKKKKKKKKKKKK EEEEE

           (R,K){20}
79: EEEEE KKKKKKKRKRKRKKKKRKRK RKRKK

           (R,K){20}
80: EEEEK KKKKKKRKRKRKKKKRKRKR KRKKK

           (R,K){20}
81: EEEKK KKKKKRKRKRKKKKRKRKRK RKKKK

           (R,K){20}
82: EEKKK KKKKRKRKRKKKKRKRKRKR KKKKK

           (R,K){20}
83: EKKKK KKKRKRKRKKKKRKRKRKRK KKKKK

           (R,K){20}
84: KKKKK KKRKRKRKKKKRKRKRKRKK KKKKK

           (R,K){20}
85: KKKKK KRKRKRKKKKRKRKRKRKKK KKKKK

           (R,K){20}
86: KKKKK RKRKRKKKKRKRKRKRKKKK KKKKK

           (R,K){20}
87: KKKKR KRKRKKKKRKRKRKRKKKKK KKKKK

           (R,K){20}
88: KKKRK RKRKKKKRKRKRKRKKKKKK KKKKK

           (R,K){20}
89: KKRKR KRKKKKRKRKRKRKKKKKKK KKKKE

           (R,K){20}
90: KRKRK RKKKKRKRKRKRKKKKKKKK KKKEK

           (R,K){20}
91: RKRKR KKKKRKRKRKRKKKKKKKKK KKEKK

           (R,K){20}
92: KRKRK KKKRKRKRKRKKKKKKKKKK KEKKK

           (R,K){20}
93: RKRKK KKRKRKRKRKKKKKKKKKKK EKKKK

ABG11277  ck: 8026  len: 1,080 ! Abg11277 Novel human diagnostic protein #11

1          (R,K){20,20}
            (R,K){20}
709: QKEKE KKKKKKKKKKKKKKKKRKKR KKKKK

            (R,K){20}
710: KEKEK KKKKKKKKKKKKKKKRKKRK KKKKK

            (R,K){20}
711: EKEKK KKKKKKKKKKKKKKRKKRKK KKKKK

            (R,K){20}
712: KEKKK KKKKKKKKKKKKKRKKRKKK KKKKK

            (R,K){20}
713: EKKKK KKKKKKKKKKKKRKKRKKKK KKKKK

            (R,K){20}
714: KKKKK KKKKKKKKKKKRKKRKKKKK KKKKK

            (R,K){20}
715: KKKKK KKKKKKKKKKRKKRKKKKKK KKKKK

            (R,K){20}
716: KKKKK KKKKKKKKKRKKRKKKKKKK KKKKK

            (R,K){20}
717: KKKKK KKKKKKKKRKKRKKKKKKKK KKKKK

            (R,K){20}
718: KKKKK KKKKKKKRKKRKKKKKKKKK KKKKK
```

```
               (R,K){20}
719: KKKKK KKKKKKRKKRKKKKKKKKKK KKKKK

               (R,K){20}
720: KKKKK KKKKKRKKRKKKKKKKKKKK KKKKK

               (R,K){20}
721: KKKKK KKKKRKKRKKKKKKKKKKKK KKKKK

               (R,K){20}
722: KKKKK KKKRKKRKKKKKKKKKKKKK KKKKK

               (R,K){20}
723: KKKKK KKRKKRKKKKKKKKKKKKKK KKKKE

               (R,K){20}
724: KKKKK KRKKRKKKKKKKKKKKKKKK KKKEE

               (R,K){20}
725: KKKKK RKKRKKKKKKKKKKKKKKKK KKEER

               (R,K){20}
726: KKKKR KKRKKKKKKKKKKKKKKKKK KEERR

               (R,K){20}
727: KKKRK KRKKKKKKKKKKKKKKKKKK EERRK

ABG11732  ck: 2886  len: 56   ! Abg11732 Novel human diagnostic protein #11
1
              (R,K){20,20}
                (R){20}
  8: KRRRG RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
  9: RRRGR RRRRRRRRRRRRRRRRRRRR RRRRP

                (R){20}
 10: RRGRR RRRRRRRRRRRRRRRRRRRR RRRPK

                (R){20}
 11: RGRRR RRRRRRRRRRRRRRRRRRRR RRPKQ

                (R){20}
 12: GRRRR RRRRRRRRRRRRRRRRRRRR RPKQK

                (R){20}
 13: RRRRR RRRRRRRRRRRRRRRRRRRR PKQKN

ABG11734  ck: 4548  len: 142   ! Abg11734 Novel human diagnostic protein #11
1
              (R,K){20,20}
               (R,K){20}
109: QMLSV KKKKKRKKKKKKKKKKKKKK KKKKK

               (R,K){20}
110: MLSVK KKKKRKKKKKKKKKKKKKKK KKKKK

               (R,K){20}
111: LSVKK KKKRKKKKKKKKKKKKKKKK KKKKK

               (R,K){20}
112: SVKKK KKRKKKKKKKKKKKKKKKKK KKKKK

               (R,K){20}
113: VKKKK KRKKKKKKKKKKKKKKKKKK KKKKK

               (R,K){20}
114: KKKKK RKKKKKKKKKKKKKKKKKKK KKKKL

                (K){20}
115: KKKKR KKKKKKKKKKKKKKKKKKKK KKKLY

                (K){20}
116: KKKRK KKKKKKKKKKKKKKKKKKKK KKLYF

                (K){20}
117: KKRKK KKKKKKKKKKKKKKKKKKKK KLYFQ

                (K){20}
118: KRKKK KKKKKKKKKKKKKKKKKKKK LYFQT

ABG11738  ck: 3009  len: 567   ! Abg11738 Novel human diagnostic protein #11
1
              (R,K){20,20}
               (R,K){20}
526: EEEEE RRKKKKKKKKKKKKKKKKKK KKKKK

               (R,K){20}
527: EEEER RKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
528: EEERR KKKKKKKKKKKKKKKKKKKK KKKKR

                (K){20}
529: EERRK KKKKKKKKKKKKKKKKKKKK KKKRK

                (K){20}
530: ERRKK KKKKKKKKKKKKKKKKKKKK KKRKE

                (K){20}
531: RRKKK KKKKKKKKKKKKKKKKKKKK KRKER

                (K){20}
532: RKKKK KKKKKKKKKKKKKKKKKKKK RKERR

               (R,K){20}
533: KKKKK KKKKKKKKKKKKKKKKKKKR KERRT

               (R,K){20}
534: KKKKK KKKKKKKKKKKKKKKKKKRK ERRTA

ABG22512  ck: 8641  len: 856   ! Abg22512 Novel human diagnostic protein #22
1
              (R,K){20,20}
                (K){20}
130: KKKKF KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
131: KKKFK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
132: KKFKK KKKKKKKKKKKKKKKKKKKK KKKKQ

                (K){20}
133: KFKKK KKKKKKKKKKKKKKKKKKKK KKKQL

                (K){20}
134: FKKKK KKKKKKKKKKKKKKKKKKKK KKQLS

                (K){20}
135: KKKKK KKKKKKKKKKKKKKKKKKKK KQLSL

                (K){20}
136: KKKKK KKKKKKKKKKKKKKKKKKKK QLSLV

ABG22638  ck: 4197  len: 896   ! Abg22638 Novel human diagnostic protein #22
1
              (R,K){20,20}
               (R,K){20}
227: GRERE RRRRRKRKRRRKRRRKRRRR RRRRG
```

```
                  (R,K){20}
       228: RERER RRRRKRKRRRKRRRKRRRRR RRRGX

                  (R,K){20}
       229: ERERR RRRKRKRRRKRRRKRRRRRR RRGXE

                  (R,K){20}
       230: RERRR RRKRKRRRKRRRKRRRRRRR RGXEF

                  (R,K){20}
       231: ERRRR RKRKRRRKRRRKRRRRRRRR GXEFL

 ABG26213  ck: 6773  len: 735   ! Abg26213 Novel human diagnostic protein #26

1                (R,K){20,20}
                  (R,K){20}
       173: RGSSS KKKKRRKKRKKKKRKKRKRR KNRKK

                  (R,K){20}
       174: GSSSK KKKRRKKRKKKKRKKRKRRK NRKKK

 ABG26488  ck: 523   len: 124   ! Abg26488 Novel human diagnostic protein #26

1                (R,K){20,20}
                  (R,K){20}
        91: EEEEE RRRRRKKKKKKKKKKKKKKK RKKKK

                  (R,K){20}
        92: EEEER RRRRKKKKKKKKKKKKKKKR KKKKK

                  (R,K){20}
        93: EEERR RRRKKKKKKKKKKKKKKKRK KKKKR

                  (R,K){20}
        94: EERRR RRKKKKKKKKKKKKKKKRKK KKKRK

                  (R,K){20}
        95: ERRRR RKKKKKKKKKKKKKKKRKKK KKRKK

                  (R,K){20}
        96: RRRRR KKKKKKKKKKKKKKKRKKKK KRKKK

                  (R,K){20}
        97: RRRRK KKKKKKKKKKKKKKRKKKKK RKKKK

                  (R,K){20}
        98: RRRKK KKKKKKKKKKKKKRKKKKKR KKKKK

                  (R,K){20}
        99: RRKKK KKKKKKKKKKKKRKKKKKRK KKKKK

                  (R,K){20}
       100: RKKKK KKKKKKKKKKKRKKKKKRKK KKKKE

                  (R,K){20}
       101: KKKKK KKKKKKKKKKRKKKKKRKKK KKKE

                  (R,K){20}
       102: KKKKK KKKKKKKKKRKKKKKRKKKK KKE

                  (R,K){20}
       103: KKKKK KKKKKKKKRKKKKKRKKKKK KE

                  (R,K){20}
       104: KKKKK KKKKKKKRKKKKKRKKKKKK E

 ABG26489  ck: 9923  len: 120   ! Abg26489 Novel human diagnostic protein #26

1                (R,K){20,20}
                   (R){20}
```

```
        70: RRRRG RRRRRRRRRRRRRRRRRRRR RRKKK

                   (R){20}
        71: RRRGR RRRRRRRRRRRRRRRRRRRR RKKKE

                   (R){20}
        72: RRGRR RRRRRRRRRRRRRRRRRRRR KKKEE

                  (R,K){20}
        73: RGRRR RRRRRRRRRRRRRRRRRRRK KKEEE

                  (R,K){20}
        74: GRRRR RRRRRRRRRRRRRRRRRRKK KEEEE

                  (R,K){20}
        75: RRRRR RRRRRRRRRRRRRRRRRKKK EEEEE

 ABG26490  ck: 4146  len: 96    ! Abg26490 Novel human diagnostic protein #26

1                (R,K){20,20}
                   (R){20}
        39: KKGEE RRRRRRRRRRRRRRRRRRRR RRRGR

                   (R){20}
        40: KGEER RRRRRRRRRRRRRRRRRRRR RRGRG

                   (R){20}
        41: GEERR RRRRRRRRRRRRRRRRRRRR RGRGR

                   (R){20}
        42: EERRR RRRRRRRRRRRRRRRRRRRR GRGRG

                   (R){20}
        73: GRGRG RRRRRRRRRRRRRRRRRRRR RRRR

                   (R){20}
        74: RGRGR RRRRRRRRRRRRRRRRRRRR RRR

                   (R){20}
        75: GRGRR RRRRRRRRRRRRRRRRRRRR RR

                   (R){20}
        76: RGRRR RRRRRRRRRRRRRRRRRRRR R

                   (R){20}
        77: GRRRR RRRRRRRRRRRRRRRRRRRR

 ABG26491  ck: 8179  len: 109   ! Abg26491 Novel human diagnostic protein #26

1                (R,K){20,20}
                  (R,K){20}
        75: RRRGE KKKKRKKKKKKKKKKKKKKK KKKKK

                  (R,K){20}
        76: RRGEK KKKRKKKKKKKKKKKKKKKK KKKKK

                  (R,K){20}
        77: RGEKK KKRKKKKKKKKKKKKKKKKK KKKKK

                  (R,K){20}
        78: GEKKK KRKKKKKKKKKKKKKKKKKK KKKKK

                  (R,K){20}
        79: EKKKK RKKKKKKKKKKKKKKKKKKK KKKKG

                   (K){20}
        80: KKKKR KKKKKKKKKKKKKKKKKKKK KKKGN

                   (K){20}
        81: KKKRK KKKKKKKKKKKKKKKKKKKK KKGNL
```

```
             (K){20}
82: KKRKK KKKKKKKKKKKKKKKKKKKK KGNLS

             (K){20}
83: KRKKK KKKKKKKKKKKKKKKKKKKK GNLSV

ABG26492   ck: 5234   len: 68    ! Abg26492 Novel human diagnostic protein #26

1          (R,K){20,20}
             (R){20}
29: KKKEE RRRRRRRRRRRRRRRRRRRR RRRRG

             (R){20}
30: KKEER RRRRRRRRRRRRRRRRRRRR RRRGK

             (R){20}
31: KEERR RRRRRRRRRRRRRRRRRRRR RRGKK

             (R){20}
32: EERRR RRRRRRRRRRRRRRRRRRRR RGKKD

             (R){20}
33: ERRRR RRRRRRRRRRRRRRRRRRRR GKKDG

ABG26493   ck: 4204   len: 80    ! Abg26493 Novel human diagnostic protein #26

1          (R,K){20,20}
            (R,K){20}
27: EKEKE KRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
28: KEKEK RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
29: EKEKR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
30: KEKRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
31: EKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
32: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
33: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
34: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
35: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
36: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
37: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
38: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
39: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
40: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
41: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
46: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
47: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

             (R){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

             (R){20}
51: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

             (R){20}
52: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

             (R){20}
53: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

ABG26496   ck: 8829   len: 90    ! Abg26496 Novel human diagnostic protein #26

1          (R,K){20,20}
            (R,K){20}
61: EKKEE KRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
62: KKEEK RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
63: KEEKR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
64: EEKRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
65: EKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
66: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

             (R){20}
67: RRRRR RRRRRRRRRRRRRRRRRRRR RRRN

             (R){20}
68: RRRRR RRRRRRRRRRRRRRRRRRRR RRN

             (R){20}
69: RRRRR RRRRRRRRRRRRRRRRRRRR RN

             (R){20}
70: RRRRR RRRRRRRRRRRRRRRRRRRR N

ABG26497   ck: 9186   len: 115   ! Abg26497 Novel human diagnostic protein #26
```

```
1                (R,K){20,20}
                   (R){20}
       24: EGEED RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       25: GEEDR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       26: EEDRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       27: EDRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       28: DRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       29: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       30: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       31: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       32: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       33: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       34: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       35: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       36: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       37: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       38: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       39: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       40: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       41: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       43: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       44: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       45: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       46: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       47: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       60: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       61: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       62: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       63: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       64: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       65: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       66: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       67: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       68: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       69: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       70: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                   (R){20}
       71: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
```

```
              (R){20}
  72: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  73: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  74: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  75: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  76: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  77: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  78: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  79: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  80: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  81: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  82: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRT

              (R){20}
  83: RRRRR RRRRRRRRRRRRRRRRRRRR RRRTT

              (R){20}
  84: RRRRR RRRRRRRRRRRRRRRRRRRR RRTTT

              (R){20}
  85: RRRRR RRRRRRRRRRRRRRRRRRRR RTTTT

              (R){20}
  86: RRRRR RRRRRRRRRRRRRRRRRRRR TTTTT

  ABG26498   ck: 7156   len: 140    ! Abg26498 Novel human diagnostic protein #26

1            (R,K){20,20}
              (R){20}
  97: RRGGG RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  98: RGGGR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  99: GGGRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 100: GGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 101: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 102: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 103: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 104: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 105: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 106: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 107: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
 108: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK

              (R){20}
 109: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKY

              (R){20}
 110: RRRRR RRRRRRRRRRRRRRRRRRRR RRKYK

              (R){20}
 111: RRRRR RRRRRRRRRRRRRRRRRRRR RKYKY

              (R){20}
 112: RRRRR RRRRRRRRRRRRRRRRRRRR KYKYL

             (R,K){20}
 113: RRRRR RRRRRRRRRRRRRRRRRRRK YKYLE

  ABG26500   ck: 7252   len: 78    ! Abg26500 Novel human diagnostic protein #26

1            (R,K){20,20}
              (R){20}
  35: GGEEE RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  36: GEEER RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  37: EEERR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  38: EERRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  39: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  40: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  41: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  43: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
  44: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRG

              (R){20}
  45: RRRRR RRRRRRRRRRRRRRRRRRRR RRRGR

              (R){20}
  46: RRRRR RRRRRRRRRRRRRRRRRRRR RRGRR

              (R){20}
  47: RRRRR RRRRRRRRRRRRRRRRRRRR RGRRR

              (R){20}
  48: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR
```

```
ABG26501   ck: 1730   len: 182    ! Abg26501 Novel human diagnostic protein #26
1                  (R,K){20,20}
                     (K){20}
       72: KEEKE KKKKKKKKKKKKKKKKKKKK KKKKE

                     (K){20}
       73: EEKEK KKKKKKKKKKKKKKKKKKKK KKKEE

                     (K){20}
       74: EKEKK KKKKKKKKKKKKKKKKKKKK KKEEE

                     (K){20}
       75: KEKKK KKKKKKKKKKKKKKKKKKKK KEEEE

                     (K){20}
       76: EKKKK KKKKKKKKKKKKKKKKKKKK EEEEA

                     (R){20}
      133: GRRRS RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      134: RRRSR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      135: RRSRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      136: RSRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      137: SRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      138: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      139: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      140: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      141: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      142: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      143: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      144: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
      145: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK

                     (R){20}
      146: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKY

                     (R){20}
      147: RRRRR RRRRRRRRRRRRRRRRRRRR RRKYP

                     (R){20}
      148: RRRRR RRRRRRRRRRRRRRRRRRRR RKYPM

                     (R){20}
      149: RRRRR RRRRRRRRRRRRRRRRRRRR KYPMW

                   (R,K){20}
      150: RRRRR RRRRRRRRRRRRRRRRRRRK YPMWV
```

```
ABG26502   ck: 3399   len: 101    ! Abg26502 Novel human diagnostic protein #26
1                  (R,K){20,20}
                   (R,K){20}
       59: EEEEG RRRRRRKKKKKKKKKKKKKK KKKKR

                   (R,K){20}
       60: EEEGR RRRRRKKKKKKKKKKKKKKK KKKRK

                   (R,K){20}
       61: EEGRR RRRRKKKKKKKKKKKKKKKK KKRKK

                   (R,K){20}
       62: EGRRR RRRKKKKKKKKKKKKKKKKK KRKKK

                   (R,K){20}
       63: GRRRR RRKKKKKKKKKKKKKKKKKK RKKKK

                   (R,K){20}
       64: RRRRR RKKKKKKKKKKKKKKKKKKR KKKKK

                   (R,K){20}
       65: RRRRR KKKKKKKKKKKKKKKKKKRK KKKKE

                   (R,K){20}
       66: RRRRK KKKKKKKKKKKKKKKKKRKK KKKEY

                   (R,K){20}
       67: RRRKK KKKKKKKKKKKKKKKKRKKK KKEYF

                   (R,K){20}
       68: RRKKK KKKKKKKKKKKKKKKRKKKK KEYFQ

                   (R,K){20}
       69: RKKKK KKKKKKKKKKKKKKRKKKKK EYFQT

ABG26505   ck: 4704   len: 93    ! Abg26505 Novel human diagnostic protein #26
1                  (R,K){20,20}
                     (R){20}
       60: KEEEG RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
       61: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
       62: EEGRR RRRRRRRRRRRRRRRRRRRR RRRRK

                     (R){20}
       63: EGRRR RRRRRRRRRRRRRRRRRRRR RRRKK

                     (R){20}
       64: GRRRR RRRRRRRRRRRRRRRRRRRR RRKKK

                     (R){20}
       65: RRRRR RRRRRRRRRRRRRRRRRRRR RKKKE

                     (R){20}
       66: RRRRR RRRRRRRRRRRRRRRRRRRR KKKEE

                   (R,K){20}
       67: RRRRR RRRRRRRRRRRRRRRRRRRK KKEEE

                   (R,K){20}
       68: RRRRR RRRRRRRRRRRRRRRRRRKK KEEEE

                   (R,K){20}
       69: RRRRR RRRRRRRRRRRRRRRRRKKK EEEEE

ABG26506   ck: 3684   len: 85    ! Abg26506 Novel human diagnostic protein #26
```

```
1                 (R,K){20,20}
                   (R,K){20}
        47: EKEKE KRKRKRKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        48: KEKEK RKRKRKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        49: EKEKR KRKRKKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        50: KEKRK RKRKKKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        51: EKRKR KRKKKKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        52: KRKRK RKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        53: RKRKR KKKKKKKKKKKKKKKKKKKK KKKKR

                    (K){20}
        54: KRKRK KKKKKKKKKKKKKKKKKKKK KKKRK

                    (K){20}
        55: RKRKK KKKKKKKKKKKKKKKKKKKK KKRKK

                    (K){20}
        56: KRKKK KKKKKKKKKKKKKKKKKKKK KRKKK

                    (K){20}
        57: RKKKK KKKKKKKKKKKKKKKKKKKK RKKKK

                   (R,K){20}
        58: KKKKK KKKKKKKKKKKKKKKKKKKR KKKKN

                   (R,K){20}
        59: KKKKK KKKKKKKKKKKKKKKKKKRK KKKNW

                   (R,K){20}
        60: KKKKK KKKKKKKKKKKKKKKKKRKK KKNWT

                   (R,K){20}
        61: KKKKK KKKKKKKKKKKKKKKKRKKK KNWTV

                   (R,K){20}
        62: KKKKK KKKKKKKKKKKKKKKRKKKK NWTV

        ABG26507   ck: 9838   len: 109   ! Abg26507 Novel human diagnostic protein #26

1                 (R,K){20,20}
                    (R){20}
        60: EEEEE RRRRRRRRRRRRRRRRRRRR RRRKK

                    (R){20}
        61: EEEER RRRRRRRRRRRRRRRRRRRR RRKKE

                    (R){20}
        62: EEERR RRRRRRRRRRRRRRRRRRRR RKKER

                    (R){20}
        63: EERRR RRRRRRRRRRRRRRRRRRRR KKERK

                   (R,K){20}
        64: ERRRR RRRRRRRRRRRRRRRRRRRK KERKK

                   (R,K){20}
        65: RRRRR RRRRRRRRRRRRRRRRRRKK ERKKE

        ABG26508   ck: 4488   len: 121   ! Abg26508 Novel human diagnostic protein #26

1                 (R,K){20,20}
                   (R,K){20}
        56: DDEEE RRRKKKKKKKKKKKKKKRKKK KKKRK

                   (R,K){20}
        57: DEEER RRKKKKKKKKKKKKKRKKKK KKRKK

                   (R,K){20}
        58: EEERR RKKKKKKKKKKKKKRKKKKK KRKKK

                   (R,K){20}
        59: EERRR KKKKKKKKKKKKKRKKKKKK RKKKK

                   (R,K){20}
        60: ERRRK KKKKKKKKKKKKRKKKKKKR KKKKK

                   (R,K){20}
        61: RRRKK KKKKKKKKKKKRKKKKKKRK KKKKK

                   (R,K){20}
        62: RRKKK KKKKKKKKKKRKKKKKKRKK KKKKK

                   (R,K){20}
        63: RKKKK KKKKKKKKKRKKKKKKRKKK KKKKK

                   (R,K){20}
        64: KKKKK KKKKKKKKRKKKKKKRKKKK KKKKK

                   (R,K){20}
        65: KKKKK KKKKKKKRKKKKKKRKKKKK KKKKN

                   (R,K){20}
        66: KKKKK KKKKKKRKKKKKKRKKKKKK KKKNI

                   (R,K){20}
        67: KKKKK KKKKKRKKKKKKRKKKKKKK KKNIL

                   (R,K){20}
        68: KKKKK KKKKRKKKKKKRKKKKKKKK KNILY

                   (R,K){20}
        69: KKKKK KKKRKKKKKKRKKKKKKKKK NILYR

        ABG26510   ck: 3426   len: 74    ! Abg26510 Novel human diagnostic protein #26

1                 (R,K){20,20}
                    (R){20}
        47: RRRRS RRRRRRRRRRRRRRRRRRRR RRRRN

                    (R){20}
        48: RRRSR RRRRRRRRRRRRRRRRRRRR RRRNT

                    (R){20}
        49: RRSRR RRRRRRRRRRRRRRRRRRRR RRNTE

                    (R){20}
        50: RSRRR RRRRRRRRRRRRRRRRRRRR RNTEK

                    (R){20}
        51: SRRRR RRRRRRRRRRRRRRRRRRRR NTEK

        ABG26513   ck: 3117   len: 265   ! Abg26513 Novel human diagnostic protein #26

1                 (R,K){20,20}
                   (R,K){20}
       199: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

                    (K){20}
       222: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE
```

```
                (K){20}
223: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

               (R,K){20}
224: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

               (R,K){20}
225: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

 ABG26514  ck: 9106  len: 218   ! Abg26514 Novel human diagnostic protein #26

1              (R,K){20,20}
                (R){20}
182: EERRG RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
183: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
184: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
185: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
186: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
187: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK

                (R){20}
188: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKK

                (R){20}
189: RRRRR RRRRRRRRRRRRRRRRRRRR RRKKE

                (R){20}
190: RRRRR RRRRRRRRRRRRRRRRRRRR RKKER

                (R){20}
191: RRRRR RRRRRRRRRRRRRRRRRRRR KKERI

               (R,K){20}
192: RRRRR RRRRRRRRRRRRRRRRRRRK KERIS

               (R,K){20}
193: RRRRR RRRRRRRRRRRRRRRRRRKK ERISS

 ABG26515  ck: 7288  len: 389   ! Abg26515 Novel human diagnostic protein #26

1              (R,K){20,20}
               (R,K){20}
259: KKEES RKKKKKKKKKKKKKKKKKKK KKKRK

                (K){20}
260: KEESR KKKKKKKKKKKKKKKKKKKK KKRKR

                (K){20}
261: EESRK KKKKKKKKKKKKKKKKKKKK KRKRN

                (K){20}
262: ESRKK KKKKKKKKKKKKKKKKKKKK RKRNK

               (R,K){20}
263: SRKKK KKKKKKKKKKKKKKKKKKKR KRNKN

               (R,K){20}
264: RKKKK KKKKKKKKKKKKKKKKKKRK RNKNK

               (R,K){20}
265: KKKKK KKKKKKKKKKKKKKKKKRKR NKNKK

 ABG26516  ck: 2295  len: 91   ! Abg26516 Novel human diagnostic protein #265

1              (R,K){20,20}
               (R,K){20}
45: EEEEE RRRRRRRRRRRRRRRRRKKR KRKRK

               (R,K){20}
46: EEEER RRRRRRRRRRRRRRRRKKRK RKRKR

               (R,K){20}
47: EEERR RRRRRRRRRRRRRRRKKRKR KRKRK

               (R,K){20}
48: EERRR RRRRRRRRRRRRRRKKRKRK RKRKK

               (R,K){20}
49: ERRRR RRRRRRRRRRRRRKKRKRKR KRKKK

               (R,K){20}
50: RRRRR RRRRRRRRRRRRKKRKRKRK RKKKT

               (R,K){20}
51: RRRRR RRRRRRRRRRRKKRKRKRKR KKKTK

               (R,K){20}
52: RRRRR RRRRRRRRRRKKRKRKRKRK KKTKT

               (R,K){20}
53: RRRRR RRRRRRRRRKKRKRKRKRKK KTKTK

               (R,K){20}
54: RRRRR RRRRRRRRKKRKRKRKRKKK TKTKT

 ABG26518  ck: 8431  len: 761   ! Abg26518 Novel human diagnostic protein #265

1              (R,K){20,20}
                (R){20}
725: EEEEE RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
726: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
727: EEERR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
728: EERRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
729: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
730: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
731: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK

                (R){20}
732: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKK

                (R){20}
733: RRRRR RRRRRRRRRRRRRRRRRRRR RRKKK

                (R){20}
734: RRRRR RRRRRRRRRRRRRRRRRRRR RKKKR

                (R){20}
735: RRRRR RRRRRRRRRRRRRRRRRRRR KKKRK
```

```
         (R,K){20}
736: RRRRR RRRRRRRRRRRRRRRRRRRK KKRKK

         (R,K){20}
737: RRRRR RRRRRRRRRRRRRRRRRRKK KRKKN

         (R,K){20}
738: RRRRR RRRRRRRRRRRRRRRRRKKK RKKN

         (R,K){20}
739: RRRRR RRRRRRRRRRRRRRRRKKKR KKN

         (R,K){20}
740: RRRRR RRRRRRRRRRRRRRRKKKRK KN

         (R,K){20}
741: RRRRR RRRRRRRRRRRRRRKKKRKK N

 ABG26520  ck: 2487  len: 99    ! Abg26520 Novel human diagnostic protein #26

1       (R,K){20,20}
          (R){20}
 35: EEEEE RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 36: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 37: EEERR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 38: EERRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 39: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 40: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 41: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 43: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 44: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 45: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 46: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 47: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 60: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 61: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 62: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 63: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 64: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 65: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 66: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 67: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 68: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 69: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 70: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 71: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 72: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 73: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
 74: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK

          (R){20}
 75: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKI

          (R){20}
```

```
 76: RRRRR RRRRRRRRRRRRRRRRRRRR RRKI

           (R){20}
 77: RRRRR RRRRRRRRRRRRRRRRRRRR RKI

           (R){20}
 78: RRRRR RRRRRRRRRRRRRRRRRRRR KI

          (R,K){20}
 79: RRRRR RRRRRRRRRRRRRRRRRRRK I

ABG26521  ck: 482   len: 367   ! Abg26521 Novel human diagnostic protein #26
1         (R,K){20,20}
           (K){20}
317: EEEEG KKKKKKKKKKKKKKKKKKKK KKKKR

           (K){20}
318: EEEGK KKKKKKKKKKKKKKKKKKKK KKKRR

           (K){20}
319: EEGKK KKKKKKKKKKKKKKKKKKKK KKRRR

           (K){20}
320: EGKKK KKKKKKKKKKKKKKKKKKKK KRRRK

           (K){20}
321: GKKKK KKKKKKKKKKKKKKKKKKKK RRRKK

          (R,K){20}
322: KKKKK KKKKKKKKKKKKKKKKKKKR RRKKK

          (R,K){20}
323: KKKKK KKKKKKKKKKKKKKKKKKRR RKKKK

          (R,K){20}
324: KKKKK KKKKKKKKKKKKKKKKKRRR KKKKE

          (R,K){20}
325: KKKKK KKKKKKKKKKKKKKKKRRRK KKKEQ

          (R,K){20}
326: KKKKK KKKKKKKKKKKKKKKRRRKK KKEQQ

          (R,K){20}
327: KKKKK KKKKKKKKKKKKKKRRRKKK KEQQL

          (R,K){20}
328: KKKKK KKKKKKKKKKKKKRRRKKKK EQQLE

ABG26522  ck: 2060  len: 152   ! Abg26522 Novel human diagnostic protein #26
1         (R,K){20,20}
           (R){20}
 46: EEEEE RRRRRRRRRRRRRRRRRRRR RRKKE

           (R){20}
 47: EEEER RRRRRRRRRRRRRRRRRRRR RKKEG

           (R){20}
 48: EEERR RRRRRRRRRRRRRRRRRRRR KKEGE

          (R,K){20}
 49: EERRR RRRRRRRRRRRRRRRRRRRK KEGEE

          (R,K){20}
 50: ERRRR RRRRRRRRRRRRRRRRRRKK EGEEG

           (R){20}
 96: KKEEG RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
 97: KEEGR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
 98: EEGRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
 99: EGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
100: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
101: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
102: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
103: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
104: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
105: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
106: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
107: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
108: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK

           (R){20}
109: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKE

           (R){20}
110: RRRRR RRRRRRRRRRRRRRRRRRRR RRKER

           (R){20}
111: RRRRR RRRRRRRRRRRRRRRRRRRR RKERE

           (R){20}
112: RRRRR RRRRRRRRRRRRRRRRRRRR KERER

          (R,K){20}
113: RRRRR RRRRRRRRRRRRRRRRRRRK ERERE

ABG26525  ck: 5433  len: 103   ! Abg26525 Novel human diagnostic protein #26
1         (R,K){20,20}
          (R,K){20}
  3:    RD RRRKKKKKKKKKKKKKKKKK EEEEE

ABG26526  ck: 1887  len: 115   ! Abg26526 Novel human diagnostic protein #26
1         (R,K){20,20}
           (R){20}
 48: EEEEE RRRRRRRRRRRRRRRRRRRR KKREK

          (R,K){20}
 49: EEEER RRRRRRRRRRRRRRRRRRRK KREKK

          (R,K){20}
 50: EEERR RRRRRRRRRRRRRRRRRRKK REKKK

          (R,K){20}
 51: EERRR RRRRRRRRRRRRRRRRRKKR EKKKS
```

```
     ABG26527  ck: 5586  len: 122   ! Abg26527 Novel human diagnostic protein #26
1               (R,K){20,20}
                 (R,K){20}
     82: EEEEE KKKKKKKKKKKKKRRRRRR GRRRM

     ABG26528  ck: 2237  len: 215   ! Abg26528 Novel human diagnostic protein #26
1               (R,K){20,20}
                 (R,K){20}
     87: EKEKE KKRKKKRKKKKKKKKKKKKK KKKKK

                 (R,K){20}
     88: KEKEK KRKKKRKKKKKKKKKKKKKK KKKKK

                 (R,K){20}
     89: EKEKK RKKKRKKKKKKKKKKKKKKK KKKKK

                 (R,K){20}
     90: KEKKR KKKRKKKKKKKKKKKKKKKK KKKKK

                 (R,K){20}
     91: EKKRK KKRKKKKKKKKKKKKKKKKK KKKKK

                 (R,K){20}
     92: KKRKK KRKKKKKKKKKKKKKKKKKK KKKKE

                 (R,K){20}
     93: KRKKK RKKKKKKKKKKKKKKKKKKK KKKEE

                 (K){20}
     94: RKKKR KKKKKKKKKKKKKKKKKKKK KKEEE

                 (K){20}
     95: KKKRK KKKKKKKKKKKKKKKKKKKK KEEEE

                 (K){20}
     96: KKRKK KKKKKKKKKKKKKKKKKKKK EEEEE

                 (R){20}
    167: GRRRG RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    168: RRRGR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    169: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    170: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    171: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    172: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    173: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRM

                 (R){20}
    174: RRRRR RRRRRRRRRRRRRRRRRRRR RRRMR

                 (R){20}
    175: RRRRR RRRRRRRRRRRRRRRRRRRR RRMRK

                 (R){20}
    176: RRRRR RRRRRRRRRRRRRRRRRRRR RMRKR

                 (R){20}
    177: RRRRR RRRRRRRRRRRRRRRRRRRR MRKRR

     ABG26530  ck: 5729  len: 404   ! Abg26530 Novel human diagnostic protein #26
1               (R,K){20,20}
                 (R){20}
    366: EEEEE RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    367: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    368: EEERR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    369: EERRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    370: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    371: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    372: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    373: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    374: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRK

                 (R){20}
    375: RRRRR RRRRRRRRRRRRRRRRRRRR RRRKN

                 (R){20}
    376: RRRRR RRRRRRRRRRRRRRRRRRRR RRKNT

                 (R){20}
    377: RRRRR RRRRRRRRRRRRRRRRRRRR RKNTS

                 (R){20}
    378: RRRRR RRRRRRRRRRRRRRRRRRRR KNTSL

                (R,K){20}
    379: RRRRR RRRRRRRRRRRRRRRRRRRK NTSLR

     ABG26531  ck: 7434  len: 126   ! Abg26531 Novel human diagnostic protein #26
1               (R,K){20,20}
                 (R){20}
     48: EEKDE RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     49: EKDER RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     50: KDERR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     51: DERRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     52: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRE

                 (R){20}
     54: RRRRR RRRRRRRRRRRRRRRRRRRR RRREE
```

```
            (R){20}
55: RRRRR RRRRRRRRRRRRRRRRRRRR RREEE

            (R){20}
56: RRRRR RRRRRRRRRRRRRRRRRRRR REEEE

            (R){20}
57: RRRRR RRRRRRRRRRRRRRRRRRRR EEEEE

ABG26532   ck: 9650   len: 225    ! Abg26532 Novel human diagnostic protein #26

1          (R,K){20,20}
            (R,K){20}
152: EEEEE KKKRRRRRRRRRRRRRRRRR RRRRR

            (R,K){20}
153: EEEEK KKRRRRRRRRRRRRRRRRRR RRRRR

            (R,K){20}
154: EEEKK KRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
155: EEKKK RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
156: EKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
157: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
158: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
159: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
160: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
161: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
162: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
163: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

            (R){20}
164: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

            (R){20}
165: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

            (R){20}
166: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

            (R){20}
167: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

ABG26533   ck: 5900   len: 101    ! Abg26533 Novel human diagnostic protein #26

1          (R,K){20,20}
            (R){20}
51: RRRRT RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
52: RRRTR RRRRRRRRRRRRRRRRRRRR RRRRK

            (R){20}
53: RRTRR RRRRRRRRRRRRRRRRRRRR RRRKK

            (R){20}
54: RTRRR RRRRRRRRRRRRRRRRRRRR RRKKK

            (R){20}
55: TRRRR RRRRRRRRRRRRRRRRRRRR RKKKP

            (R){20}
56: RRRRR RRRRRRRRRRRRRRRRRRRR KKKPT

            (R,K){20}
57: RRRRR RRRRRRRRRRRRRRRRRRRK KKPTS

            (R,K){20}
58: RRRRR RRRRRRRRRRRRRRRRRRKK KPTSR

            (R,K){20}
59: RRRRR RRRRRRRRRRRRRRRRRKKK PTSRV

ABG26534   ck: 1919   len: 113    ! Abg26534 Novel human diagnostic protein #26

1          (R,K){20,20}
            (R,K){20}
31: KKKKE KKRKKRRKRKKKKKKKKKKK KKKKK

            (R,K){20}
32: KKKEK KRKKRRKRKKKKKKKKKKKK KKKKK

            (R,K){20}
33: KKEKK RKKRRKRKKKKKKKKKKKKK KKKKK

            (R,K){20}
34: KEKKR KKRRKRKKKKKKKKKKKKKK KKKKK

            (R,K){20}
35: EKKRK KRRKRKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
36: KKRKK RRKRKKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
37: KRKKR RKRKKKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
38: RKKRR KRKKKKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
39: KKRRK RKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
40: KRRKR KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
41: RRKRK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
42: RKRKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
43: KRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
44: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

            (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRK

            (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKRKK
```

```
            (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KRKKR

            (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK RKKRK

           (R,K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKR KKRKK

           (R,K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKRK KRKKQ

           (R,K){20}
52: KKKKK KKKKKKKKKKKKKKKKKRKK RKKQK

           (R,K){20}
53: KKKKK KKKKKKKKKKKKKKKKRKKR KKQKK

           (R,K){20}
54: KKKKK KKKKKKKKKKKKKKKRKKRK KQKKR

           (R,K){20}
55: KKKKK KKKKKKKKKKKKKKRKKRKK QKKRR

ABG26535  ck: 2214  len: 472   ! Abg26535 Novel human diagnostic protein #26

1         (R,K){20,20}
            (R){20}
415: EEEEE RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
416: EEEER RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
417: EEERR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
418: EERRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
419: ERRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
420: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
421: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
422: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
423: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
424: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
425: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
426: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
427: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
428: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
429: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
430: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
431: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
432: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
433: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
434: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRG

            (R){20}
435: RRRRR RRRRRRRRRRRRRRRRRRRR RRRGG

            (R){20}
436: RRRRR RRRRRRRRRRRRRRRRRRRR RRGGR

            (R){20}
437: RRRRR RRRRRRRRRRRRRRRRRRRR RGGRR

            (R){20}
438: RRRRR RRRRRRRRRRRRRRRRRRRR GGRRR

ABG26537  ck: 2429  len: 573   ! Abg26537 Novel human diagnostic protein #26

1         (R,K){20,20}
            (R){20}
405: GAEQE RRRRRRRRRRRRRRRRRRRR RREEE

            (R){20}
406: AEQER RRRRRRRRRRRRRRRRRRRR REEEE

            (R){20}
407: EQERR RRRRRRRRRRRRRRRRRRRR EEEEE

           (R,K){20}
479: KKEEE KKRKKKKKKKKKKKKKKKKK KKKKK

           (R,K){20}
480: KEEEK KRKKKKKKKKKKKKKKKKKK KKKKK

           (R,K){20}
481: EEEKK RKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
482: EEKKR KKKKKKKKKKKKKKKKKKKK KKKKT

            (K){20}
483: EKKRK KKKKKKKKKKKKKKKKKKKK KKKTR

            (K){20}
484: KKRKK KKKKKKKKKKKKKKKKKKKK KKTRK

            (K){20}
485: KRKKK KKKKKKKKKKKKKKKKKKKK KTRKK

            (K){20}
486: RKKKK KKKKKKKKKKKKKKKKKKKK TRKKK

ABG26538  ck: 5732  len: 98    ! Abg26538 Novel human diagnostic protein #26

1         (R,K){20,20}
           (R,K){20}
60: KEKET KRRRRRKKKKKRKKKKKKKK KKKKK

           (R,K){20}
```

```
61: EKETK RRRRRKKKKRKKKKKKKKK KKKKK

              (R,K){20}
62: KETKR RRRRKKKKRKKKKKKKKKK KKKKK

              (R,K){20}
63: ETKRR RRRKKKKRKKKKKKKKKKK KKKKK

              (R,K){20}
64: TKRRR RRKKKKRKKKKKKKKKKKK KKKKK

              (R,K){20}
65: KRRRR RKKKKRKKKKKKKKKKKKK KKKKK

              (R,K){20}
66: RRRRR KKKKRKKKKKKKKKKKKKK KKKKK

              (R,K){20}
67: RRRRK KKKRKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
68: RRRKK KKRKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
69: RRKKK KRKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
70: RKKKK KRKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
71: KKKKK RKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
72: KKKKR KKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
73: KKKRK KKKKKKKKKKKKKKKKKKK KKKKG

               (K){20}
74: KKRKK KKKKKKKKKKKKKKKKKKK KKKGY

               (K){20}
75: KRKKK KKKKKKKKKKKKKKKKKKK KKGY

               (K){20}
76: RKKKK KKKKKKKKKKKKKKKKKKK KGY

               (K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKK GY

ABG26539   ck: 3545   len: 89     ! Abg26539 Novel human diagnostic protein #26

1            (R,K){20,20}
               (R){20}
28: EEEEE RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
29: EEEER RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
30: EEERR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
31: EERRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
32: ERRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
33: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
34: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
35: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
36: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
37: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
38: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
39: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
40: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
41: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

               (R){20}
46: RRRRR RRRRRRRRRRRRRRRRRRR RRRRS

               (R){20}
47: RRRRR RRRRRRRRRRRRRRRRRRR RRRSH

               (R){20}
48: RRRRR RRRRRRRRRRRRRRRRRRR RRSHT

               (R){20}
49: RRRRR RRRRRRRRRRRRRRRRRRR RSHTS

               (R){20}
50: RRRRR RRRRRRRRRRRRRRRRRRR SHTSI

ABG26542   ck: 4605   len: 182     ! Abg26542 Novel human diagnostic protein #265

1            (R,K){20,20}
               (R){20}
146: EGEEE RRRRRRRRRRRRRRRRRRR RRRGR

               (R){20}
147: GEEER RRRRRRRRRRRRRRRRRRR RRGRR

               (R){20}
148: EEERR RRRRRRRRRRRRRRRRRRR RGRRG

               (R){20}
149: EERRR RRRRRRRRRRRRRRRRRRR GRRGR

ABG26543   ck: 2998   len: 405     ! Abg26543 Novel human diagnostic protein #265

1            (R,K){20,20}
              (R,K){20}
264: RRRRW RRRRRRRRRKKKKKKKKKK KKKKK
```

```
         (R,K){20}
265: RRRWR RRRRRRRRKKKKKKKKKKKK KKKKK

         (R,K){20}
266: RRWRR RRRRRRRKKKKKKKKKKKKK KKKKK

         (R,K){20}
267: RWRRR RRRRRRKKKKKKKKKKKKKK KKKKK

         (R,K){20}
268: WRRRR RRRRRKKKKKKKKKKKKKKK KKKKK

         (R,K){20}
269: RRRRR RRRRKKKKKKKKKKKKKKKK KKKKK

         (R,K){20}
270: RRRRR RRRKKKKKKKKKKKKKKKKK KKKKK

         (R,K){20}
271: RRRRR RRKKKKKKKKKKKKKKKKKK KKKKK

         (R,K){20}
272: RRRRR RKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
273: RRRRR KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
274: RRRRK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
275: RRRKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
276: RRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
277: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
278: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

          (K){20}
279: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGK

          (K){20}
280: KKKKK KKKKKKKKKKKKKKKKKKKK KKGKK

          (K){20}
281: KKKKK KKKKKKKKKKKKKKKKKKKK KGKKG

          (K){20}
282: KKKKK KKKKKKKKKKKKKKKKKKKK GKKGK

 ABG26545  ck: 135  len: 119  ! Abg26545 Novel human diagnostic protein #26

1        (R,K){20,20}
          (K){20}
 88: EKEKE KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 89: KEKEK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 90: EKEKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 91: KEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 92: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
 93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKN

          (K){20}
 95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKNF

          (K){20}
 96: KKKKK KKKKKKKKKKKKKKKKKKKK KKNF

          (K){20}
 97: KKKKK KKKKKKKKKKKKKKKKKKKK KNF

          (K){20}
 98: KKKKK KKKKKKKKKKKKKKKKKKKK NF

ABG26551  ck: 8976  len: 94   ! Abg26551 Novel human diagnostic protein #26

1         (R,K){20,20}
          (R,K){20}
  1:       KRRRKKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
  2:     K RRRKKKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
  3:    KR RRKKKKKKKKKKKKKKKKKK KKKKR

          (R,K){20}
  4:   KRR RKKKKKKKKKKKKKKKKKKK KKKRK

          (K){20}
  5:  KRRR KKKKKKKKKKKKKKKKKKKK KKRKK

          (K){20}
  6: KRRRK KKKKKKKKKKKKKKKKKKKK KRKKK

          (K){20}
  7: RRRKK KKKKKKKKKKKKKKKKKKKK RKKKR

          (R,K){20}
  8: RRKKK KKKKKKKKKKKKKKKKKKKR KKKRK

          (R,K){20}
  9: RKKKK KKKKKKKKKKKKKKKKKKRK KKRKK

          (R,K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKRKK KRKKK

          (R,K){20}
 11: KKKKK KKKKKKKKKKKKKKKKRKKK RKKKR

          (R,K){20}
 12: KKKKK KKKKKKKKKKKKKKKRKKKR KKKRK

          (R,K){20}
 13: KKKKK KKKKKKKKKKKKKKRKKKRK KKRKK

          (R,K){20}
 14: KKKKK KKKKKKKKKKKKKRKKKRKK KRKKK

          (R,K){20}
 15: KKKKK KKKKKKKKKKKKRKKKRKKK RKKKK

          (R,K){20}
 16: KKKKK KKKKKKKKKKKRKKKRKKKR KKKKK

          (R,K){20}
 17: KKKKK KKKKKKKKKKRKKKRKKKRK KKKKE
```

```
          (R,K){20}
18: KKKKK KKKKKKKKKRKKKRKKKRKK KKKEK

          (R,K){20}
19: KKKKK KKKKKKKKRKKKRKKKRKKK KKEKK

          (R,K){20}
20: KKKKK KKKKKKKRKKKRKKKRKKKK KEKKK

          (R,K){20}
21: KKKKK KKKKKKRKKKRKKKRKKKKK EKKKK

ABG26717  ck: 4415  len: 78    ! Abg26717 Novel human diagnostic protein #26

1        (R,K){20,20}
          (R,K){20}
37: QMLSV KKKKKRKKKKKKKKKKKKKK KKKKK

          (R,K){20}
38: MLSVK KKKKRKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
39: LSVKK KKKRKKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
40: SVKKK KKRKKKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
41: VKKKK KRKKKKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
42: KKKKK RKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
43: KKKKR KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
44: KKKRK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
45: KKRKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
46: KRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
47: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKL

           (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKLY

           (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKLYF

           (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KLYFQ

           (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK LYFQT

ABG26718  ck: 9531  len: 141   ! Abg26718 Novel human diagnostic protein #26

1        (R,K){20,20}
          (R,K){20}
80: EVARP RKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
81: VARPR KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
82: ARPRK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
83: RPRKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
84: PRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
85: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

           (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

           (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

           (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

           (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEE

ABG26719  ck: 72    len: 83    ! Abg26719 Novel human diagnostic protein #26

1        (R,K){20,20}
           (K){20}
42: ETPSE KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
43: TPSEK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
44: PSEKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
45: SEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
46: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKN

           (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKNP

           (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKNPI

           (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KNPIF

           (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK NPIFF
```

```
ABG26720   ck: 3781   len: 57    ! Abg26720 Novel human diagnostic protein #26
1               (R,K){20,20}
                  (K){20}
24: QMKSI KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
25: MKSIK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
26: KSIKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
27: SIKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
28: IKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKF

                  (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKFH

                  (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKFHF

                  (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KFHFL

                  (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK FHFLL

ABG26721   ck: 287   len: 95    ! Abg26721 Novel human diagnostic protein #26
1               (R,K){20,20}
                  (K){20}
31: EKEKE KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
32: KEKEK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
33: EKEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
34: KEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
35: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKN

                  (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKNK

                  (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKNKK

                  (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KNKKK

                  (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK NKKKK

ABG26722   ck: 4831   len: 127    ! Abg26722 Novel human diagnostic protein #26
1               (R,K){20,20}
                  (K){20}
44: KKKEE KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
45: KKEEK KKKKKKKKKKKKKKKKKKKK KKKKX

                  (K){20}
46: KEEKK KKKKKKKKKKKKKKKKKKKK KKKXE

                  (K){20}
47: EEKKK KKKKKKKKKKKKKKKKKKKK KKXEK

                  (K){20}
48: EKKKK KKKKKKKKKKKKKKKKKKKK KXEKE

                  (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK XEKEK

                  (K){20}
87: EKEEE KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
88: KEEEK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
89: EEEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
90: EEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
```

```
91: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK
          (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKK KKK
          (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKK KK
          (K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKKK K
          (K){20}
108: KKKKK KKKKKKKKKKKKKKKKKKKK

ABG26723  ck: 7054  len: 98   ! Abg26723 Novel human diagnostic protein #26

1         (R,K){20,20}
          (K){20}
42: EKEKE KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
43: KEKEK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
44: EKEKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
45: KEKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
46: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKQ
          (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKQL
          (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK KKQLL
          (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK KQLLS
          (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKK QLLSK

ABG26724  ck: 9897  len: 295   ! Abg26724 Novel human diagnostic protein #26
```

```
1                 (R,K){20,20}
                   (K){20}
        259: KEEEE KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        260: EEEEK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        261: EEEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        262: EEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        263: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        264: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        265: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        266: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        267: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        268: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        269: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        270: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        271: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        272: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                   (K){20}
        273: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                   (K){20}
        274: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                   (K){20}
        275: KKKKK KKKKKKKKKKKKKKKKKKKK K

                   (K){20}
        276: KKKKK KKKKKKKKKKKKKKKKKKKK

         ABG26725   ck: 909   len: 131   ! Abg26725 Novel human diagnostic protein #26

1                 (R,K){20,20}
                   (R,K){20}
         70: KKKKE KKRKKKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
         71: KKKEK KRKKKKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
         72: KKEKK RKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         73: KEKKR KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         74: EKKRK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         75: KKRKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         76: KRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         77: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         78: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         79: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKM

                   (K){20}
         80: KKKKK KKKKKKKKKKKKKKKKKKKK KKKML

                   (K){20}
         81: KKKKK KKKKKKKKKKKKKKKKKKKK KKMLL

                   (K){20}
         82: KKKKK KKKKKKKKKKKKKKKKKKKK KMLLL

                   (K){20}
         83: KKKKK KKKKKKKKKKKKKKKKKKKK MLLLS

         ABG26726   ck: 9300   len: 677   ! Abg26726 Novel human diagnostic protein #26

1                 (R,K){20,20}
                   (R,K){20}
        240: EKEKE RKRKRKRKRKKKKKKKKKKK KKKKK

                   (R,K){20}
        241: KEKER KRKRKRKRKKKKKKKKKKKK KKKKK

                   (R,K){20}
        242: EKERK RKRKRKRKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        243: KERKR KRKRKRKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        244: ERKRK RKRKRKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        245: RKRKR KRKRKKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        246: KRKRK RKRKKKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        247: RKRKR KRKKKKKKKKKKKKKKKKKK KKKKK

                   (R,K){20}
        248: KRKRK RKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        249: RKRKR KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        250: KRKRK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        251: RKRKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        252: KRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
```

```
253: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
254: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
255: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
256: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR
                (K){20}
257: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRK
                (K){20}
258: KKKKK KKKKKKKKKKKKKKKKKKKK KKRKK
                (K){20}
259: KKKKK KKKKKKKKKKKKKKKKKKKK KRKKK
                (K){20}
260: KKKKK KKKKKKKKKKKKKKKKKKKK RKKKK
              (R,K){20}
261: KKKKK KKKKKKKKKKKKKKKKKKKR KKKKL
              (R,K){20}
262: KKKKK KKKKKKKKKKKKKKKKKKRK KKKLS
              (R,K){20}
263: KKKKK KKKKKKKKKKKKKKKKKRKK KKLSL
              (R,K){20}
264: KKKKK KKKKKKKKKKKKKKKKRKKK KLSLH
              (R,K){20}
265: KKKKK KKKKKKKKKKKKKKKRKKKK LSLHL

 ABG26727  ck: 813   len: 329   ! Abg26727 Novel human diagnostic protein #26
1             (R,K){20,20}
                (K){20}
215: EKEKE KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
216: KEKEK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
217: EKEKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
218: KEKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
219: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
220: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
221: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
222: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
223: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
224: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
225: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
226: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKT
                (K){20}
227: KKKKK KKKKKKKKKKKKKKKKKKKK KKKTS
                (K){20}
228: KKKKK KKKKKKKKKKKKKKKKKKKK KKTSQ
                (K){20}
229: KKKKK KKKKKKKKKKKKKKKKKKKK KTSQD
                (K){20}
230: KKKKK KKKKKKKKKKKKKKKKKKKK TSQDA

 ABG26879  ck: 2669  len: 800   ! Abg26879 Novel human diagnostic protein #26
1             (R,K){20,20}
               (R,K){20}
559: EREQE KKRKKKKKKKKKKKKKKKKK KKKKK
               (R,K){20}
560: REQEK KRKKKKKKKKKKKKKKKKKK KKKKQ
               (R,K){20}
561: EQEKK RKKKKKKKKKKKKKKKKKKK KKKQQ
                (K){20}
562: QEKKR KKKKKKKKKKKKKKKKKKKK KKQQE
                (K){20}
563: EKKRK KKKKKKKKKKKKKKKKKKKK KQQEK
                (K){20}
564: KKRKK KKKKKKKKKKKKKKKKKKKK QQEKE

 ABG28885  ck: 8668  len: 137   ! Abg28885 Novel human diagnostic protein #28
1             (R,K){20,20}
                (R){20}
 35: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRK
                (R){20}
 36: EEEGR RRRRRRRRRRRRRRRRRRRR RRRKK
                (R){20}
 37: EEGRR RRRRRRRRRRRRRRRRRRRR RRKKV
                (R){20}
 38: EGRRR RRRRRRRRRRRRRRRRRRRR RKKVL
                (R){20}
 39: GRRRR RRRRRRRRRRRRRRRRRRRR KKVLS
               (R,K){20}
 40: RRRRR RRRRRRRRRRRRRRRRRRRK KVLSP
               (R,K){20}
 41: RRRRR RRRRRRRRRRRRRRRRRRKK VLSPS

 AAG65985  ck: 8085  len: 154   ! Aag65985 B726P splice variant sequence. 2/2
1             (R,K){20,20}
                (K){20}
114: TQLRQ KKKKKKKKKKKKKKKKKKKK KKKKK
                (K){20}
115: QLRQK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
116: LRQKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
117: RQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
118: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
121: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
122: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
123: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
124: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
125: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                (K){20}
126: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

                (K){20}
127: KKKKK KKKKKKKKKKKKKKKKKKKK KXKKK

                (K){20}
128: KKKKK KKKKKKKKKKKKKKKKKKKK XKKKK

 ABB27893   ck: 5383   len: 86    ! Abb27893 Human peptide #544 encoded by brea

1              (R,K){20,20}
                (R,K){20}
 15: RRRRG RRRRRKKKKKKRKKKKKRKK KKRRR

                (R,K){20}
 16: RRRGR RRRRKKKKKKRKKKKKRKKK KRRRR

                (R,K){20}
 17: RRGRR RRRKKKKKKRKKKKKRKKKK RRRRR

                (R,K){20}
 18: RGRRR RRKKKKKKRKKKKKRKKKKR RRRRR

                (R,K){20}
 19: GRRRR RKKKKKKRKKKKKRKKKKRR RRRRR

                (R,K){20}
 20: RRRRR KKKKKKRKKKKKRKKKKRRR RRRRR

                (R,K){20}
 21: RRRRK KKKKKRKKKKKRKKKKRRRR RRRRR

                (R,K){20}
 22: RRRKK KKKKRKKKKKRKKKKRRRRR RRRRR

                (R,K){20}
 23: RRKKK KKKRKKKKKRKKKKRRRRRR RRRRR

                (R,K){20}
 24: RKKKK KKRKKKKKRKKKKRRRRRRR RRRRR

                (R,K){20}
 25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

                (R,K){20}
 26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

                (R,K){20}
 27: KKKKR KKKKRKKKKRRRRRRRRRRR RRRRR

                (R,K){20}
 28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

                (R,K){20}
 29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

                (R,K){20}
 30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

                (R,K){20}
 31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

                (R,K){20}
 32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
 33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
 34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
 35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
 36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
 41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                 (R){20}
 42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

                 (R){20}
 43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

                 (R){20}
 44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

                 (R){20}
 45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

 ABB28750   ck: 1334   len: 86    ! Abb28750 Peptide #1401 encoded by breast ce

1              (R,K){20,20}
                (R,K){20}
 57: EEEEG RRRKKKKKRKRKKKKKKKKK KKKKK

                (R,K){20}
 58: EEEGR RRKKKKKRKRKKKKKKKKKK KKKKK
```

```
          (R,K){20}
59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

          (R,K){20}
60: EGRRR KKKKRKRKKKKKKKKKKKK KKKKK

          (R,K){20}
61: GRRRK KKKRKRKKKKKKKKKKKKK KKKKK

          (R,K){20}
62: RRRKK KKRKRKKKKKKKKKKKKKK KKKKK

          (R,K){20}
63: RRKKK KRKRKKKKKKKKKKKKKKK KKKK

          (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKK KKK

          (R,K){20}
65: KKKKR KRKKKKKKKKKKKKKKKKK KK

          (R,K){20}
66: KKKRK RKKKKKKKKKKKKKKKKKK K

           (K){20}
67: KKRKR KKKKKKKKKKKKKKKKKKKK

ABB28840  ck: 9082  len: 167   ! Abb28840 Peptide #1491 encoded by breast ce

1        (R,K){20,20}
          (R,K){20}
33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

          (R,K){20}
34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

          (R,K){20}
35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR

ABB29360  ck: 2276  len: 89    ! Abb29360 Peptide #2011 encoded by breast ce

1        (R,K){20,20}
          (R,K){20}
23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

           (K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

           (K){20}
47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

          (R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

          (R,K){20}
49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

ABB29645  ck: 1939  len: 130   ! Abb29645 Peptide #2296 encoded by breast ce

1        (R,K){20,20}
          (R,K){20}
42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR

          (R,K){20}
43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

          (R,K){20}
44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK
```

```
          (R,K){20}
45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

          (R,K){20}
46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

          (R,K){20}
47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

          (R,K){20}
48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

          (R,K){20}
49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

          (R,K){20}
50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

          (R,K){20}
51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

          (R,K){20}
52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

          (R,K){20}
53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE

          (R,K){20}
54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE

ABB29928  ck: 3607  len: 88    ! Abb29928 Peptide #2579 encoded by breast ce

1        (R,K){20,20}
           (K){20}
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
             (K){20}
 52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

             (K){20}
 66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

             (K){20}
 67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

             (K){20}
 68: KKKKK KKKKKKKKKKKKKKKKKKKK K

             (K){20}
 69: KKKKK KKKKKKKKKKKKKKKKKKKK

ABB30512  ck: 3937  len: 85   ! Abb30512 Peptide #3163 encoded by breast ce

1           (R,K){20,20}
             (K){20}
  1:       KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
  7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

             (K){20}
  8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

             (K){20}
  9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

             (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

             (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

ABB32308  ck: 1560  len: 88   ! Abb32308 Peptide #4959 encoded by breast ce

1           (R,K){20,20}
             (R){20}
 43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

             (R){20}
 59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT
```

```
              (R){20}
60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

              (R){20}
61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

              (R){20}
62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

ABB33064  ck: 5383  len: 86   ! Abb33064 Peptide #570 encoded by human foet
1
            (R,K){20,20}
             (R,K){20}
15: RRRRG RRRRRKKKKKKRKKKKKRKK KKRRR

             (R,K){20}
16: RRRGR RRRRKKKKKKRKKKKKRKKK KRRRR

             (R,K){20}
17: RRGRR RRRKKKKKKRKKKKKRKKKK RRRRR

             (R,K){20}
18: RGRRR RRKKKKKKRKKKKKRKKKKR RRRRR

             (R,K){20}
19: GRRRR RKKKKKKRKKKKKRKKKKRR RRRRR

             (R,K){20}
20: RRRRR KKKKKKRKKKKKRKKKKRRR RRRRR

             (R,K){20}
21: RRRRK KKKKKRKKKKKRKKKKRRRR RRRRR

             (R,K){20}
22: RRRKK KKKKRKKKKKRKKKKRRRRR RRRRR

             (R,K){20}
23: RRKKK KKKRKKKKKRKKKKRRRRRR RRRRR

             (R,K){20}
24: RKKKK KKRKKKKKRKKKKRRRRRRR RRRRR

             (R,K){20}
25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

             (R,K){20}
26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

             (R,K){20}
27: KKKKR KKKKKRKKKKRRRRRRRRRR RRRRR

             (R,K){20}
28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

             (R,K){20}
29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

             (R,K){20}
30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

             (R,K){20}
31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

             (R,K){20}
32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

             (R,K){20}
33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

             (R,K){20}
34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

             (R,K){20}
35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

             (R,K){20}
36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

              (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

              (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

              (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

              (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

ABB33937  ck: 1334  len: 86   ! Abb33937 Peptide #1443 encoded by human foet
1
            (R,K){20,20}
             (R,K){20}
57: EEEEG RRRKKKKRKRKKKKKKKKKK KKKKK

             (R,K){20}
58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

             (R,K){20}
59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

             (R,K){20}
60: EGRRR KKKKRKRKKKKKKKKKKKKK KKKKK

             (R,K){20}
61: GRRRK KKKRKRKKKKKKKKKKKKKK KKKKK

             (R,K){20}
62: RRRKK KKRKRKKKKKKKKKKKKKKK KKKKK

             (R,K){20}
63: RRKKK KRKRKKKKKKKKKKKKKKKK KKKK

             (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKKK KKK

             (R,K){20}
65: KKKKR KRKKKKKKKKKKKKKKKKKK KK

             (R,K){20}
66: KKKRK RKKKKKKKKKKKKKKKKKKK K

              (K){20}
67: KKRKR KKKKKKKKKKKKKKKKKKKK

ABB34024  ck: 9082  len: 167   ! Abb34024 Peptide #1530 encoded by human foet
```

```
1                (R,K){20,20}
                  (R,K){20}
         33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

                  (R,K){20}
         34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

                  (R,K){20}
         35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR


         ABB34533  ck: 2276  len: 89   ! Abb34533 Peptide #2039 encoded by human foe

1                (R,K){20,20}
                  (R,K){20}
         23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

                   (K){20}
         46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

                   (K){20}
         47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

                  (R,K){20}
         48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

                  (R,K){20}
         49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE


         ABB34819  ck: 1939  len: 130   ! Abb34819 Peptide #2325 encoded by human foe

1                (R,K){20,20}
                  (R,K){20}
         42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR

                  (R,K){20}
         43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

                  (R,K){20}
         44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK

                  (R,K){20}
         45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

                  (R,K){20}
         46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

                  (R,K){20}
         47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

                  (R,K){20}
         48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

                  (R,K){20}
         49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

                  (R,K){20}
         50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

                  (R,K){20}
         51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

                  (R,K){20}
         52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

                  (R,K){20}
         53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE

                  (R,K){20}
         54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE


         ABB35110  ck: 3607  len: 88   ! Abb35110 Peptide #2616 encoded by human foe

1                (R,K){20,20}
                   (K){20}
         39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
         61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
```

```
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

            (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

            (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

            (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK K

            (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK

ABB35676   ck: 3937   len: 85    ! Abb35676 Peptide #3182 encoded by human foe

1           (R,K){20,20}
            (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

            (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

            (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

            (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

ABB36406   ck: 2686   len: 71    ! Abb36406 Peptide #3912 encoded by human foe

1           (R,K){20,20}
            (K){20}
20: KKKKQ KKKKKKKKKKKKKKKKKKKK KKKKR

            (K){20}
21: KKKQK KKKKKKKKKKKKKKKKKKKK KKKRK

            (K){20}
22: KKQKK KKKKKKKKKKKKKKKKKKKK KKRKR

            (K){20}
23: KQKKK KKKKKKKKKKKKKKKKKKKK KRKRK

            (K){20}
24: QKKKK KKKKKKKKKKKKKKKKKKKK RKRKK

           (R,K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKR KRKKK

           (R,K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKRK RKKKK

           (R,K){20}
27: KKKKK KKKKKKKKKKKKKKKKKRKR KKKKK

           (R,K){20}
28: KKKKK KKKKKKKKKKKKKKKKRKRK KKKKK

           (R,K){20}
29: KKKKK KKKKKKKKKKKKKKKRKRKK KKKKK

           (R,K){20}
30: KKKKK KKKKKKKKKKKKKKRKRKKK KKKKK

           (R,K){20}
31: KKKKK KKKKKKKKKKKKKRKRKKKK KKKKK

           (R,K){20}
32: KKKKK KKKKKKKKKKKKRKRKKKKK KKKKK

           (R,K){20}
33: KKKKK KKKKKKKKKKKRKRKKKKKK KKKKK

           (R,K){20}
34: KKKKK KKKKKKKKKKRKRKKKKKKK KKKKK

           (R,K){20}
35: KKKKK KKKKKKKKKRKRKKKKKKKK KKKKK

           (R,K){20}
36: KKKKK KKKKKKKKRKRKKKKKKKKK KKKKK

           (R,K){20}
37: KKKKK KKKKKKKRKRKKKKKKKKKK KKKKK

           (R,K){20}
38: KKKKK KKKKKKRKRKKKKKKKKKKK KKKKK

           (R,K){20}
39: KKKKK KKKKKRKRKKKKKKKKKKKK KKKKK

           (R,K){20}
40: KKKKK KKKKRKRKKKKKKKKKKKKK KKKKK

           (R,K){20}
41: KKKKK KKKRKRKKKKKKKKKKKKKK KKKKK

           (R,K){20}
42: KKKKK KKRKRKKKKKKKKKKKKKKK KKKKK

           (R,K){20}
43: KKKKK KRKRKKKKKKKKKKKKKKKK KKKKK

           (R,K){20}
44: KKKKK RKRKKKKKKKKKKKKKKKKK KKKKK

           (R,K){20}
45: KKKKR KRKKKKKKKKKKKKKKKKKK KKKKS

           (R,K){20}
46: KKKRK RKKKKKKKKKKKKKKKKKKK KKKSA

            (K){20}
```

```
         47: KKRKR KKKKKKKKKKKKKKKKKKKK KKSAH

                    (K){20}
         48: KRKRK KKKKKKKKKKKKKKKKKKKK KSAH

                    (K){20}
         49: RKRKK KKKKKKKKKKKKKKKKKKKK SAH

         ABB37567   ck: 1560  len: 88    ! Abb37567 Peptide #5073 encoded by human foe

1                  (R,K){20,20}
                    (R){20}
         43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                    (R){20}
         58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                    (R){20}
         59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

                    (R){20}
         60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

                    (R){20}
         61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

                    (R){20}
         62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE
```

```
         ABB37780   ck: 2324  len: 36    ! Abb37780 Peptide #5286 encoded by human foe

1                  (R,K){20,20}
                    (R,K){20}
         16: KERKT KKRKRKRRRRRKKRRRKKRR R

                    (R,K){20}
         17: ERKTK KRKRKRRRRRKKRRRKKRRR

         ABB40272   ck: 8343  len: 66    ! Abb40272 Peptide #7778 encoded by human foe

1                  (R,K){20,20}
                    (R,K){20}
          6: ETERE KKKKRKKKKKKKKKKKKKKK KKKNK

                    (R,K){20}
          7: TEREK KKKRKKKKKKKKKKKKKKKK KKNKK

                    (R,K){20}
          8: EREKK KKRKKKKKKKKKKKKKKKKK KNKKK

                    (R,K){20}
          9: REKKK KRKKKKKKKKKKKKKKKKKK NKKKK

         ABB42642   ck: 2394  len: 57    ! Abb42642 Peptide #10148 encoded by human fc

1                  (R,K){20,20}
                     (R){20}
         20: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRR

                     (R){20}
         21: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRG

                     (R){20}
         22: EEGRR RRRRRRRRRRRRRRRRRRRR RRRGR

                     (R){20}
         23: EGRRR RRRRRRRRRRRRRRRRRRRR RRGRR

                     (R){20}
         24: GRRRR RRRRRRRRRRRRRRRRRRRR RGRRR

                     (R){20}
         25: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR

         ABB43181   ck: 4228  len: 24    ! Abb43181 Peptide #10687 encoded by human fc

1                  (R,K){20,20}
                    (R,K){20}
          1:       RRRRRRRKKKKRRRRRRRRR RKRT

                    (R,K){20}
          2:     R RRRRRRKKKKRRRRRRRRRR KRT

                    (R,K){20}
          3:    RR RRRRRKKKKRRRRRRRRRRK RT

                    (R,K){20}
          4:   RRR RRRRKKKKRRRRRRRRRRKR T

         ABB44317   ck: 4895  len: 51    ! Abb44317 Peptide #11823 encoded by human fo

1                  (R,K){20,20}
                    (R,K){20}
         18: LFKPM RKRRRKKKRRRKKRKKKRRR KKLTT

                    (R,K){20}
         19: FKPMR KRRRKKKRRRKKRKKKRRRK KLTTT
```

```
                (R,K){20}
20: KPMRK RRRKKKRRRKKRKKKRRRKK LTTTT

ABB17165  ck: 8887  len: 42    ! Abb17165 Human nervous system related polyp

1              (R,K){20,20}
                (K){20}
21: FFTTE KKKKKKKKKKKKKKKKKKKK KX

                (K){20}
22: FTTEK KKKKKKKKKKKKKKKKKKKK X

ABB18534  ck: 5383  len: 86    ! Abb18534 Protein #533 encoded by probe for

1              (R,K){20,20}
                (R,K){20}
15: RRRRG RRRRRKKKKKRKKKKKRKK KKRRR

                (R,K){20}
16: RRRGR RRRRKKKKKKRKKKKKRKKK KRRRR

                (R,K){20}
17: RRGRR RRRKKKKKKRKKKKKRKKKK RRRRR

                (R,K){20}
18: RGRRR RRKKKKKKRKKKKKRKKKKR RRRRR

                (R,K){20}
19: GRRRR RKKKKKKRKKKKKRKKKKRR RRRRR

                (R,K){20}
20: RRRRR KKKKKKRKKKKKRKKKKRRR RRRRR

                (R,K){20}
21: RRRRK KKKKKRKKKKKRKKKKRRRR RRRRR

                (R,K){20}
22: RRRKK KKKKRKKKKKRKKKKRRRRR RRRRR

                (R,K){20}
23: RRKKK KKKRKKKKKRKKKKRRRRRR RRRRR

                (R,K){20}
24: RKKKK KKRKKKKKRKKKKRRRRRRR RRRRR

                (R,K){20}
25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

                (R,K){20}
26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

                (R,K){20}
27: KKKKR KKKKKRKKKKRRRRRRRRRR RRRRR

                (R,K){20}
28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

                (R,K){20}
29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

                (R,K){20}
30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

                (R,K){20}
31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

                (R,K){20}
32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                 (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

                 (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

                 (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

                 (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

ABB19373  ck: 1334  len: 86    ! Abb19373 Protein #1372 encoded by probe for

1              (R,K){20,20}
                (R,K){20}
57: EEEEG RRRKKKKRKRKKKKKKKKKK KKKKK

                (R,K){20}
58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

                (R,K){20}
59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

                (R,K){20}
60: EGRRR KKKKRKRKKKKKKKKKKKKK KKKKK

                (R,K){20}
61: GRRRK KKKRKRKKKKKKKKKKKKKK KKKKK

                (R,K){20}
62: RRRKK KKRKRKKKKKKKKKKKKKKK KKKKK

                (R,K){20}
63: RRKKK KRKRKKKKKKKKKKKKKKKK KKKK

                (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKKK KKK

                (R,K){20}
65: KKKKR KRKKKKKKKKKKKKKKKKKK KK

                (R,K){20}
66: KKKRK RKKKKKKKKKKKKKKKKKKK K

                 (K){20}
67: KKRKR KKKKKKKKKKKKKKKKKKKK
```

```
     ABB19467  ck: 9082  len: 167   ! Abb19467 Protein #1466 encoded by probe for

1              (R,K){20,20}
                (R,K){20}
     33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

                (R,K){20}
     34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

                (R,K){20}
     35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR

     ABB19943  ck: 2276  len: 89    ! Abb19943 Protein #1942 encoded by probe for

1              (R,K){20,20}
                (R,K){20}
     23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

                 (K){20}
     46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

                 (K){20}
     47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

                (R,K){20}
     48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

                (R,K){20}
     49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

     ABB20238  ck: 1939  len: 130   ! Abb20238 Protein #2237 encoded by probe for

1              (R,K){20,20}
                (R,K){20}
     42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR

                (R,K){20}
     43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

                (R,K){20}
     44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK

                (R,K){20}
     45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

                (R,K){20}
     46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

                (R,K){20}
     47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

                (R,K){20}
     48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

                (R,K){20}
     49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

                (R,K){20}
     50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

                (R,K){20}
     51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

                (R,K){20}
     52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

                (R,K){20}
     53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE

                (R,K){20}
     54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE

     ABB20531  ck: 3607  len: 88    ! Abb20531 Protein #2530 encoded by probe for

1              (R,K){20,20}
                 (K){20}
     39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK

ABB21105   ck: 3937   len: 85    ! Abb21105 Protein #3104 encoded by probe for

1           (R,K){20,20}
              (K){20}
 1:         KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

              (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

              (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

              (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

              (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

ABB21763   ck: 2686   len: 71    ! Abb21763 Protein #3762 encoded by probe for

1           (R,K){20,20}
              (K){20}
20: KKKKQ KKKKKKKKKKKKKKKKKKKK KKKKR

              (K){20}
21: KKKQK KKKKKKKKKKKKKKKKKKKK KKKRK

              (K){20}
22: KKQKK KKKKKKKKKKKKKKKKKKKK KKRKR

              (K){20}
23: KQKKK KKKKKKKKKKKKKKKKKKKK KRKRK

              (K){20}
24: QKKKK KKKKKKKKKKKKKKKKKKKK RKRKK

            (R,K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKR KRKKK

            (R,K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKRK RKKKK

            (R,K){20}
27: KKKKK KKKKKKKKKKKKKKKKKRKR KKKKK

            (R,K){20}
28: KKKKK KKKKKKKKKKKKKKKKKRKRK KKKKK

            (R,K){20}
29: KKKKK KKKKKKKKKKKKKKKRKRKK KKKKK

            (R,K){20}
30: KKKKK KKKKKKKKKKKKKKRKRKKK KKKKK

            (R,K){20}
31: KKKKK KKKKKKKKKKKKKRKRKKKK KKKKK

            (R,K){20}
32: KKKKK KKKKKKKKKKKKRKRKKKKK KKKKK

            (R,K){20}
33: KKKKK KKKKKKKKKKKRKRKKKKKK KKKKK

            (R,K){20}
34: KKKKK KKKKKKKKKKRKRKKKKKKK KKKKK

            (R,K){20}
35: KKKKK KKKKKKKKKRKRKKKKKKKK KKKKK

            (R,K){20}
36: KKKKK KKKKKKKKRKRKKKKKKKKK KKKKK

            (R,K){20}
37: KKKKK KKKKKKKRKRKKKKKKKKKK KKKKK

            (R,K){20}
38: KKKKK KKKKKKRKRKKKKKKKKKKK KKKKK

            (R,K){20}
39: KKKKK KKKKKRKRKKKKKKKKKKKK KKKKK

            (R,K){20}
40: KKKKK KKKKRKRKKKKKKKKKKKKK KKKKK

            (R,K){20}
41: KKKKK KKKRKRKKKKKKKKKKKKKK KKKKK

            (R,K){20}
42: KKKKK KKRKRKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
43: KKKKK KRKRKKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
44: KKKKK RKRKKKKKKKKKKKKKKKKK KKKKK

            (R,K){20}
45: KKKKR KRKKKKKKKKKKKKKKKKKK KKKKS
```

```
            (R,K){20}
 46: KKKRK RKKKKKKKKKKKKKKKKKKK KKKSA

             (K){20}
 47: KKRKR KKKKKKKKKKKKKKKKKKKK KKSAH

             (K){20}
 48: KRKRK KKKKKKKKKKKKKKKKKKKK KSAH

             (K){20}
 49: RKRKK KKKKKKKKKKKKKKKKKKKK SAH

ABB22862  ck: 1560  len: 88   ! Abb22862 Protein #4861 encoded by probe for

1           (R,K){20,20}
             (R){20}
 43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

             (R){20}
 59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

             (R){20}
 60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

             (R){20}
 61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

             (R){20}
 62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

ABB23064  ck: 2324  len: 36   ! Abb23064 Protein #5063 encoded by probe for

1           (R,K){20,20}
            (R,K){20}
 16: KERKT KKRKRKRRRRRKKRRRKKRR R

            (R,K){20}
 17: ERKTK KRKRKRRRRRKKRRRKKRRR

ABB24685  ck: 8343  len: 66   ! Abb24685 Protein #6684 encoded by probe for

1           (R,K){20,20}
            (R,K){20}
  6: ETERE KKKKRKKKKKKKKKKKKKKK KKKNK

            (R,K){20}
  7: TEREK KKKRKKKKKKKKKKKKKKKK KKNKK

            (R,K){20}
  8: EREKK KKRKKKKKKKKKKKKKKKKK KNKKK

            (R,K){20}
  9: REKKK KRKKKKKKKKKKKKKKKKKK NKKKK

ABB25988  ck: 2394  len: 57   ! Abb25988 Protein #7987 encoded by probe for

1           (R,K){20,20}
             (R){20}
 20: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 21: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRG

             (R){20}
 22: EEGRR RRRRRRRRRRRRRRRRRRRR RRRGR

             (R){20}
 23: EGRRR RRRRRRRRRRRRRRRRRRRR RRGRR

             (R){20}
 24: GRRRR RRRRRRRRRRRRRRRRRRRR RGRRR

             (R){20}
 25: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR

ABB27176  ck: 4895  len: 51   ! Abb27176 Protein #9175 encoded by probe for

1           (R,K){20,20}
            (R,K){20}
 18: LFKPM RKRRRKKKRRRKKRKKKRRR KKLTT

            (R,K){20}
 19: FKPMR KRRRKKKRRRKKRKKKRRRK KLTTT

            (R,K){20}
 20: KPMRK RRRKKKRRRKKRKKKRRRKK LTTTT

ABB10296  ck: 3983  len: 292   ! Abb10296 Human cDNA SEQ ID NO: 604. 1/2002

1           (R,K){20,20}
            (R,K){20}
273: QVFAP RKKKKKKKKKKKKKKKKKKK

ABB10485  ck: 7611  len: 315   ! Abb10485 Human cDNA SEQ ID NO: 793. 1/2002
```

```
1                (R,K){20,20}
                  (R,K){20}
    273: QVFAP RKKKKKKKKKKKKKKKKKKK KGGRS

                  (K){20}
    274: VFAPR KKKKKKKKKKKKKKKKKKKK GGRSR

    AAU21948  ck: 444   len: 66    ! Aau21948 Human cardiovascular system antige

1                (R,K){20,20}
                  (K){20}
     35: SMTFS KKKKKKKKKKKKKKKKKKKK KXGKK

                  (K){20}
     36: MTFSK KKKKKKKKKKKKKKKKKKKK XGKKK

    AAU22148  ck: 4665  len: 34    ! Aau22148 Human cardiovascular system antige

1                (R,K){20,20}
                  (K){20}
     10: PELLL KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     11: ELLLK KKKKKKKKKKKKKKKKKKKK KKKK

                  (K){20}
     12: LLLKK KKKKKKKKKKKKKKKKKKKK KKK

                  (K){20}
     13: LLKKK KKKKKKKKKKKKKKKKKKKK KK

                  (K){20}
     14: LKKKK KKKKKKKKKKKKKKKKKKKK K

                  (K){20}
     15: KKKKK KKKKKKKKKKKKKKKKKKKK

    AAU22186  ck: 269   len: 76    ! Aau22186 Human cardiovascular system antige

1                (R,K){20,20}
                  (K){20}
     37: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     38: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     39: SRAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     40: RAKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                  (K){20}
     41: AKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

                  (K){20}
     42: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

                  (K){20}
     43: KKKKK KKKKKKKKKKKKKKKKKKKK KXKKI

                  (K){20}
     44: KKKKK KKKKKKKKKKKKKKKKKKKK XKKIK

    AAU22374  ck: 8278  len: 53    ! Aau22374 Human cardiovascular system antige

1                (R,K){20,20}
                  (K){20}
```

```
     30: NCGIL KKKKKKKKKKKKKKKKKKKK KKKK

                  (K){20}
     31: CGILK KKKKKKKKKKKKKKKKKKKK KKK

                  (K){20}
     32: GILKK KKKKKKKKKKKKKKKKKKKK KK

                  (K){20}
     33: ILKKK KKKKKKKKKKKKKKKKKKKK K

                  (K){20}
     34: LKKKK KKKKKKKKKKKKKKKKKKKK

    AAU23799  ck: 6158  len: 272   ! Aau23799 Novel human enzyme polypeptide #88!

1                (R,K){20,20}
                  (K){20}
    238: SFANA KKKKKKKKKKKKKKKKKKKK KKKKG

                  (K){20}
    239: FANAK KKKKKKKKKKKKKKKKKKKK KKKGR

                  (K){20}
    240: ANAKK KKKKKKKKKKKKKKKKKKKK KKGRP

                  (K){20}
    241: NAKKK KKKKKKKKKKKKKKKKKKKK KGRPX

                  (K){20}
    242: AKKKK KKKKKKKKKKKKKKKKKKKK GRPXX

    AAU27944  ck: 1121  len: 69    ! Aau27944 Human contig polypeptide sequence

1                (R,K){20,20}
                  (R,K){20}
     40: VPPLT RKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     41: PPLTR KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     42: PLTRK KKKKKKKKKKKKKKKKKKKK KKKKR

                  (K){20}
     43: LTRKK KKKKKKKKKKKKKKKKKKKK KKKRG

                  (K){20}
     44: TRKKK KKKKKKKKKKKKKKKKKKKK KKRGA

                  (K){20}
     45: RKKKK KKKKKKKKKKKKKKKKKKKK KRGAL

                  (K){20}
     46: KKKKK KKKKKKKKKKKKKKKKKKKK RGAL

                 (R,K){20}
     47: KKKKK KKKKKKKKKKKKKKKKKKKR GAL

    AAU31467  ck: 4264  len: 657   ! Aau31467 Novel human secreted protein #1958

1                (R,K){20,20}
                  (R){20}
     19: RRRRP RRRRRRRRRRRRRRRRRRRR RRRRL

                  (R){20}
     20: RRRPR RRRRRRRRRRRRRRRRRRRR RRRLG

                  (R){20}
     21: RRPRR RRRRRRRRRRRRRRRRRRRR RRLGL
```

```
              (R){20}
   22: RPRRR RRRRRRRRRRRRRRRRRRRR RLGLE

              (R){20}
   23: PRRRR RRRRRRRRRRRRRRRRRRRR LGLER

  AAU33348  ck: 8085  len: 154   ! Aau33348 Human breast cancer protein encode
1            (R,K){20,20}
              (K){20}
  114: TQLRQ KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  115: QLRQK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  116: LRQKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  117: RQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  118: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  121: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  122: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  123: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  124: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
  125: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

              (K){20}
  126: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

              (K){20}
  127: KKKKK KKKKKKKKKKKKKKKKKKKK KXKKK

              (K){20}
  128: KKKKK KKKKKKKKKKKKKKKKKKKK XKKKK

  AAE09664  ck: 1663  len: 87    ! Aae09664 Human pancreatic related protein H
1            (R,K){20,20}
              (K){20}
   36: KWSSX KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   37: WSSXK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   38: SSXKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   39: SXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   40: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

              (K){20}
   59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

              (K){20}
   60: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGX

              (K){20}
   61: KKKKK KKKKKKKKKKKKKKKKKKKK KRGXP

              (K){20}
   62: KKKKK KKKKKKKKKKKKKKKKKKKK RGXPF

             (R,K){20}
   63: KKKKK KKKKKKKKKKKKKKKKKKKR GXPFX

  AAM95365  ck: 5626  len: 139   ! Aam95365 Human reproductive system related
```

```
1              (R,K){20,20}
                 (K){20}
    111: IHLNL KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
    112: HLNLK KKKKKKKKKKKKKKKKKKKK KKXKK

                 (K){20}
    113: LNLKK KKKKKKKKKKKKKKKKKKKK KXKKK

                 (K){20}
    114: NLKKK KKKKKKKKKKKKKKKKKKKK XKKKK

     AAM96607  ck: 4751  len: 80    ! Aam96607 Human reproductive system related

1              (R,K){20,20}
                 (K){20}
     61: KKXFD KKKKKKKKKKKKKKKKKKKK

     AAU18162  ck: 7907  len: 39    ! Aau18162 Novel human DNA-binding protein #9

1              (R,K){20,20}
                 (K){20}
      9: YFEDL KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     10: FEDLK KKKKKKKKKKKKKKKKKKKK KKKKG

                 (K){20}
     11: EDLKK KKKKKKKKKKKKKKKKKKKK KKKGX

                 (K){20}
     12: DLKKK KKKKKKKKKKKKKKKKKKKK KKGXA

                 (K){20}
     13: LKKKK KKKKKKKKKKKKKKKKKKKK KGXAA

                 (K){20}
     14: KKKKK KKKKKKKKKKKKKKKKKKKK GXAAL

     AAU18167  ck: 9194  len: 87    ! Aau18167 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
     52: KIILL KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     53: IILLK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
     54: ILLKK KKKKKKKKKKKKKKKKKKKK KKKXK

                 (K){20}
     55: LLKKK KKKKKKKKKKKKKKKKKKKK KKXKK

                 (K){20}
     56: LKKKK KKKKKKKKKKKKKKKKKKKK KXKKK

                 (K){20}
     57: KKKKK KKKKKKKKKKKKKKKKKKKK XKKKK

     AAU18168  ck: 8659  len: 104   ! Aau18168 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
     75: PLGGQ KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     76: LGGQK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
     77: GGQKK KKKKKKKKKKKKKKKKKKKK KKKXG

                 (K){20}
     78: GQKKK KKKKKKKKKKKKKKKKKKKK KKXGX

                 (K){20}
     79: QKKKK KKKKKKKKKKKKKKKKKKKK KXGXP

                 (K){20}
     80: KKKKK KKKKKKKKKKKKKKKKKKKK XGXPF

     AAU18171  ck: 9398  len: 48    ! Aau18171 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
      2:     Q KKKKKKKKKKKKKKKKKKKK KKGGR

                 (K){20}
      3:    QK KKKKKKKKKKKKKKKKKKKK KGGRX

                 (K){20}
      4:   QKK KKKKKKKKKKKKKKKKKKKK GGRXR

     AAU18177  ck: 8278  len: 53    ! Aau18177 Novel human DNA-binding protein #2

1              (R,K){20,20}
                 (K){20}
     30: NCGIL KKKKKKKKKKKKKKKKKKKK KKKK

                 (K){20}
     31: CGILK KKKKKKKKKKKKKKKKKKKK KKK

                 (K){20}
     32: GILKK KKKKKKKKKKKKKKKKKKKK KK

                 (K){20}
     33: ILKKK KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
     34: LKKKK KKKKKKKKKKKKKKKKKKKK

     AAU18178  ck: 444   len: 66    ! Aau18178 Novel human DNA-binding protein #2

1              (R,K){20,20}
                 (K){20}
     35: SMTFS KKKKKKKKKKKKKKKKKKKK KXGKK

                 (K){20}
     36: MTFSK KKKKKKKKKKKKKKKKKKKK XGKKK

     AAU18179  ck: 5503  len: 50    ! Aau18179 Novel human DNA-binding protein #2

1              (R,K){20,20}
                 (K){20}
     30: IICLL KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
     31: ICLLK KKKKKKKKKKKKKKKKKKKK

     AAU18184  ck: 5691  len: 108   ! Aau18184 Novel human DNA-binding protein #3

1              (R,K){20,20}
                 (K){20}
     78: VRPCL KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
             (K){20}
79: RPCLK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
80: PCLKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
81: CLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

             (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

             (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKK KK

             (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKK K

             (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKKK

AAU18192   ck: 6029   len: 63    ! Aau18192 Novel human DNA-binding protein #3

1          (R,K){20,20}
             (K){20}
40: KLTLL KKKKKKKKKKKKKKKKKKKK ISWG

AAU18200   ck: 7170   len: 63    ! Aau18200 Novel human DNA-binding protein #4

1          (R,K){20,20}
             (K){20}
37: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
38: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
39: SRAKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
40: RAKKK KKKKKKKKKKKKKKKKKKKK KKKK

             (K){20}
41: AKKKK KKKKKKKKKKKKKKKKKKKK KKK

             (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KK

             (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK K

             (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK

AAU18204   ck: 6110   len: 61    ! Aau18204 Novel human DNA-binding protein #5

1          (R,K){20,20}
             (K){20}
```

```
28: RPTRP KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
29: PTRPK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
30: TRPKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
31: RPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
32: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

             (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

             (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

             (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGG

             (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK GGGGG

AAU18205   ck: 5764   len: 74    ! Aau18205 Novel human DNA-binding protein #5

1          (R,K){20,20}
             (K){20}
40: EFLSA KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
41: FLSAK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
42: LSAKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
43: SAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
44: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

             (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKX

             (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

             (K){20}
```

```
        53: KKKKK KKKKKKKKKKKKKKKKKKKK KX

                    (K){20}
        54: KKKKK KKKKKKKKKKKKKKKKKKKK X

        AAU18206   ck: 9217  len: 68    ! Aau18206 Novel human DNA-binding protein #5

1                 (R,K){20,20}
                    (K){20}
        38: FLFPE KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        39: LFPEK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        40: FPEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        41: PEKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                    (K){20}
        42: EKKKK KKKKKKKKKKKKKKKKKKKK KKKGX

                    (K){20}
        43: KKKKK KKKKKKKKKKKKKKKKKKKK KKGXX

                    (K){20}
        44: KKKKK KKKKKKKKKKKKKKKKKKKK KGXXF

                    (K){20}
        45: KKKKK KKKKKKKKKKKKKKKKKKKK GXXF

        AAU18208   ck: 8162  len: 79    ! Aau18208 Novel human DNA-binding protein #5

1                 (R,K){20,20}
                   (R,K){20}
        41: VRPRV RKKKKKKKKKKKKKKKKKKK KKKKG

                    (K){20}
        42: RPRVR KKKKKKKKKKKKKKKKKKKK KKKGG

                    (K){20}
        43: PRVRK KKKKKKKKKKKKKKKKKKKK KKGGR

                    (K){20}
        44: RVRKK KKKKKKKKKKKKKKKKKKKK KGGRF

                    (K){20}
        45: VRKKK KKKKKKKKKKKKKKKKKKKK GGRFR

        AAU18237   ck: 285   len: 118   ! Aau18237 Novel human DNA-binding protein #8

1                 (R,K){20,20}
                   (R,K){20}
        98: EKHKQ KKKKKKKKKKKKKKKKKKKR G

        AAU18238   ck: 5509  len: 58    ! Aau18238 Novel human DNA-binding protein #8

1                 (R,K){20,20}
                    (K){20}
        36: FYFVC KKKKKKKKKKKKKKKKKKKK KKK

                    (K){20}
        37: YFVCK KKKKKKKKKKKKKKKKKKKK KK

                    (K){20}
        38: FVCKK KKKKKKKKKKKKKKKKKKKK K

                    (K){20}
        39: VCKKK KKKKKKKKKKKKKKKKKKKK

        AAU18239   ck: 9074  len: 66    ! Aau18239 Novel human DNA-binding protein #86

1                 (R,K){20,20}
                    (K){20}
        40: LVQCE KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        41: VQCEK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        42: QCEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        43: CEKKK KKKKKKKKKKKKKKKKKKKK KKKK

                    (K){20}
        44: EKKKK KKKKKKKKKKKKKKKKKKKK KKK

                    (K){20}
        45: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                    (K){20}
        46: KKKKK KKKKKKKKKKKKKKKKKKKK K

                    (K){20}
        47: KKKKK KKKKKKKKKKKKKKKKKKKK

        AAU18240   ck: 8528  len: 150   ! Aau18240 Novel human DNA-binding protein #87

1                 (R,K){20,20}
                    (K){20}
       113: SRNTV KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       114: RNTVK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       115: NTVKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       116: TVKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       117: VKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       118: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                    (K){20}
       120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXX

                    (K){20}
       121: KKKKK KKKKKKKKKKKKKKKKKKKK KKXXA

                    (K){20}
       122: KKKKK KKKKKKKKKKKKKKKKKKKK KXXAV

                    (K){20}
       123: KKKKK KKKKKKKKKKKKKKKKKKKK XXAVL

        AAU18241   ck: 7676  len: 156   ! Aau18241 Novel human DNA-binding protein #88

1                 (R,K){20,20}
                    (K){20}
       108: KTTWI KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
     109: TTWIK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     110: TWIKK KKKKKKKKKKKKKKKKKKKK KKKKR

                (K){20}
     111: WIKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                (K){20}
     112: IKKKK KKKKKKKKKKKKKKKKKKKK KKRGX

                (K){20}
     113: KKKKK KKKKKKKKKKKKKKKKKKKK KRGXG

                (K){20}
     114: KKKKK KKKKKKKKKKKKKKKKKKKK RGXGS

               (R,K){20}
     115: KKKKK KKKKKKKKKKKKKKKKKKKR GXGST

      AAU18242  ck: 1736  len: 40    ! Aau18242 Novel human DNA-binding protein #8

1              (R,K){20,20}
                (K){20}
      18: LPGSL KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
      19: PGSLK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
      20: GSLKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
      21: SLKKK KKKKKKKKKKKKKKKKKKKK

      AAU18244  ck: 1109  len: 98    ! Aau18244 Novel human DNA-binding protein #9

1              (R,K){20,20}
                (K){20}
      53: QTKNT KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      54: TKNTK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      55: KNTKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      56: NTKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      57: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
      65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
      66: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGR

                (K){20}
      67: KKKKK KKKKKKKKKKKKKKKKKKKK KGGRS

                (K){20}
      68: KKKKK KKKKKKKKKKKKKKKKKKKK GGRSR

      AAU18246  ck: 8102  len: 111   ! Aau18246 Novel human DNA-binding protein #9

1              (R,K){20,20}
                (K){20}
      78: EFHIL KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      79: FHILK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      80: HILKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      81: ILKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      87: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
      88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
      89: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
      90: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
      91: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
      92: KKKKK KKKKKKKKKKKKKKKKKKKK

      AAU18247  ck: 8102  len: 111   ! Aau18247 Novel human DNA-binding protein #9

1              (R,K){20,20}
                (K){20}
      78: EFHIL KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
79: FHILK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
80: HILKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
81: ILKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
90: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKK

AAU18248  ck: 8319  len: 53    ! Aau18248 Novel human DNA-binding protein #9

1            (R,K){20,20}
              (K){20}
13: RYFKP KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
14: YFKPK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
15: FKPKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
16: KPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
17: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXE

              (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKXEG

              (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KXEGX

              (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK XEGX

AAU18250  ck: 7918  len: 80    ! Aau18250 Novel human DNA-binding protein #97

1            (R,K){20,20}
              (K){20}
50: NVLTV KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
51: VLTVK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
52: LTVKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
53: TVKKK KKKKKKKKKKKKKKKKKKKK KKKXX

              (K){20}
54: VKKKK KKKKKKKKKKKKKKKKKKKK KKXXK

              (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KXXKK

              (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK XXKKK

AAU18252  ck: 4882  len: 41    ! Aau18252 Novel human DNA-binding protein #9

1            (R,K){20,20}
              (K){20}
 8: FYCFF KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 9: YCFFK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
10: CFFKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
11: FFKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
12: FKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGX

              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKGX

              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KGX

              (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK GX

AAU18253  ck: 5469  len: 63    ! Aau18253 Novel human DNA-binding protein #1

1           (R,K){20,20}
              (K){20}
30: IICLL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
31: ICLLK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
32: CLLKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
33: LLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXG

              (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKXG

              (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KXG

              (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK XG

AAU18254  ck: 5075  len: 52    ! Aau18254 Novel human DNA-binding protein #1

1           (R,K){20,20}
              (K){20}
30: FIVVX KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
31: IVVXK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
32: VVXKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
33: VXKKK KKKKKKKKKKKKKKKKKKKK

AAU18255  ck: 5741  len: 47    ! Aau18255 Novel human DNA-binding protein #1

1           (R,K){20,20}
              (K){20}
20: ILTTF KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
21: LTTFK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
22: TTFKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
23: TFKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
24: FKKKK KKKKKKKKKKKKKKKKKKKK KKKX

              (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

              (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KX

              (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK X

AAU18256  ck: 2868  len: 84    ! Aau18256 Novel human DNA-binding protein #1

1           (R,K){20,20}
              (K){20}
53: KCTYE KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
54: CTYEK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
55: TYEKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
56: YEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
57: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

              (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

              (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KXKK

              (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK XKK

AAU18257  ck: 4686  len: 73    ! Aau18257 Novel human DNA-binding protein #1
```

```
1              (R,K){20,20}
                (K){20}
       41: YLKKE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       42: LKKEK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       43: KKEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       44: KEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       45: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
       51: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
       52: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
       53: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
       54: KKKKK KKKKKKKKKKKKKKKKKKKK

       AAU18258  ck: 6676  len: 74    ! Aau18258 Novel human DNA-binding protein #1

1              (R,K){20,20}
                (K){20}
       47: LRTFQ KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       48: RTFQK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
       49: TFQKK KKKKKKKKKKKKKKKKKKKK KKKXX

                (K){20}
       50: FQKKK KKKKKKKKKKKKKKKKKKKK KKXXG

                (K){20}
       51: QKKKK KKKKKKKKKKKKKKKKKKKK KXXG

                (K){20}
       52: KKKKK KKKKKKKKKKKKKKKKKKKK XXG

       AAU18259  ck: 2283  len: 54    ! Aau18259 Novel human DNA-binding protein #1

1              (R,K){20,20}
                (K){20}
       32: IVFCF KKKKKKKKKKKKKKKKKKKK KKX

                (K){20}
       33: VFCFK KKKKKKKKKKKKKKKKKKKK KX

                (K){20}
       34: FCFKK KKKKKKKKKKKKKKKKKKKK X

       AAU18260  ck: 7503  len: 74    ! Aau18260 Novel human DNA-binding protein #1(

1              (R,K){20,20}
                (K){20}
       45: SHLTD KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       46: HLTDK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       47: LTDKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       48: TDKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       49: DKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
       51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
       52: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
       53: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
       54: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
       55: KKKKK KKKKKKKKKKKKKKKKKKKK

       AAU18262  ck: 5199  len: 84    ! Aau18262 Novel human DNA-binding protein #1(

1              (R,K){20,20}
                (K){20}
       63: AWNAS KKKKKKKKKKKKKKKKKKKK XG

       AAU18263  ck: 7578  len: 31    ! Aau18263 Novel human DNA-binding protein #1

1              (R,K){20,20}
                (K){20}
        6: LTELE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
        7: TELEK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
        8: ELEKK KKKKKKKKKKKKKKKKKKKK KKKX

                (K){20}
        9: LEKKK KKKKKKKKKKKKKKKKKKKK KKX

                (K){20}
       10: EKKKK KKKKKKKKKKKKKKKKKKKK KX

                (K){20}
       11: KKKKK KKKKKKKKKKKKKKKKKKKK X

       AAU18264  ck: 3915  len: 57    ! Aau18264 Novel human DNA-binding protein #1
```

```
1              (R,K){20,20}
                 (K){20}
      31: KQLLL KKKKKKKKKKKKKKKKKKKK KKKXG

                 (K){20}
      32: QLLLK KKKKKKKKKKKKKKKKKKKK KKXGG

                 (K){20}
      33: LLLKK KKKKKKKKKKKKKKKKKKKK KXGGF

                 (K){20}
      34: LLKKK KKKKKKKKKKKKKKKKKKKK XGGF

      AAU18265   ck: 3679   len: 37    ! Aau18265 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
      15: ISPLT KKKKKKKKKKKKKKKKKKKK KKX

                 (K){20}
      16: SPLTK KKKKKKKKKKKKKKKKKKKK KX

                 (K){20}
      17: PLTKK KKKKKKKKKKKKKKKKKKKK X

      AAU18266   ck: 657    len: 196   ! Aau18266 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
     169: FVXFE KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     170: VXFEK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     171: XFEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     172: FEKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
     173: EKKKK KKKKKKKKKKKKKKKKKKKK KKKX

                 (K){20}
     174: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

                 (K){20}
     175: KKKKK KKKKKKKKKKKKKKKKKKKK KX

                 (K){20}
     176: KKKKK KKKKKKKKKKKKKKKKKKKK X

      AAU18267   ck: 4672   len: 57    ! Aau18267 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
      28: DKTFH KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
      29: KTFHK KKKKKKKKKKKKKKKKKKKK KKKXP

                 (K){20}
      30: TFHKK KKKKKKKKKKKKKKKKKKKK KKXPG

                 (K){20}
      31: FHKKK KKKKKKKKKKKKKKKKKKKK KXPGG

                 (K){20}
      32: HKKKK KKKKKKKKKKKKKKKKKKKK XPGGG

      AAU18268   ck: 9656   len: 66    ! Aau18268 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
      38: MVISV KKKKKKKKKKKKKKKKKKKK KKKKR

                 (K){20}
      39: VISVK KKKKKKKKKKKKKKKKKKKK KKKRE

                 (K){20}
      40: ISVKK KKKKKKKKKKKKKKKKKKKK KKREK

                 (K){20}
      41: SVKKK KKKKKKKKKKKKKKKKKKKK KREKX

                 (K){20}
      42: VKKKK KKKKKKKKKKKKKKKKKKKK REKXK

               (R,K){20}
      43: KKKKK KKKKKKKKKKKKKKKKKKKR EKXK

      AAU18270   ck: 4665   len: 34    ! Aau18270 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
      10: PELLL KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      11: ELLLK KKKKKKKKKKKKKKKKKKKK KKKK

                 (K){20}
      12: LLLKK KKKKKKKKKKKKKKKKKKKK KKK

                 (K){20}
      13: LLKKK KKKKKKKKKKKKKKKKKKKK KK

                 (K){20}
      14: LKKKK KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
      15: KKKKK KKKKKKKKKKKKKKKKKKKK

      AAU18271   ck: 7810   len: 64    ! Aau18271 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
      37: LKYFW KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
      38: KYFWK KKKKKKKKKKKKKKKKKKKK KKKXG

                 (K){20}
      39: YFWKK KKKKKKKKKKKKKKKKKKKK KKXGX

                 (K){20}
      40: FWKKK KKKKKKKKKKKKKKKKKKKK KXGXP

                 (K){20}
      41: WKKKK KKKKKKKKKKKKKKKKKKKK XGXP

      AAU18272   ck: 269    len: 76    ! Aau18272 Novel human DNA-binding protein #1

1              (R,K){20,20}
                 (K){20}
      37: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
      38: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
           (K){20}
39: SRAKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
40: RAKKK KKKKKKKKKKKKKKKKKKKK KKKKX

           (K){20}
41: AKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

           (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

           (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KXKKI

           (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK XKKIK

AAU18273  ck: 8370  len: 45   ! Aau18273 Novel human DNA-binding protein #1

1        (R,K){20,20}
           (K){20}
17: APKTQ KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
18: PKTQK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
19: KTQKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
20: TQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
21: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

           (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKX

           (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

           (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KX

           (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK X

AAU18274  ck: 1663  len: 87   ! Aau18274 Novel human DNA-binding protein #1

1        (R,K){20,20}
           (K){20}
36: KWSSX KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
37: WSSXK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
38: SSXKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
39: SXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
40: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

           (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

           (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGX

           (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KRGXP

           (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK RGXPF

         (R,K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKR GXPFX

AAU18275  ck: 5607  len: 63   ! Aau18275 Novel human DNA-binding protein #1

1        (R,K){20,20}
           (K){20}
26: MVELE KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
```

```
         27: VELEK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         28: ELEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         29: LEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         30: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                    (K){20}
         37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                    (K){20}
         38: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

                    (K){20}
         39: KKKKK KKKKKKKKKKKKKKKKKKKK KGGPF

                    (K){20}
         40: KKKKK KKKKKKKKKKKKKKKKKKKK GGPF

         AAU18276   ck: 5997   len: 58    ! Aau18276 Novel human DNA-binding protein #1

1                 (R,K){20,20}
                    (K){20}
         28: RPTRP KKKKKKKKKKKKKKKKKKKK KXKKK

                    (K){20}
         29: PTRPK KKKKKKKKKKKKKKKKKKKK XKKKK

         AAU18277   ck: 5764   len: 74    ! Aau18277 Novel human DNA-binding protein #1

1                 (R,K){20,20}
                    (K){20}
         40: EFLSA KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         41: FLSAK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         42: LSAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         43: SAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         44: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                    (K){20}
         51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKX

                    (K){20}
         52: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

                    (K){20}
         53: KKKKK KKKKKKKKKKKKKKKKKKKK KX

                    (K){20}
         54: KKKKK KKKKKKKKKKKKKKKKKKKK X

         AAU18278   ck: 7734   len: 97    ! Aau18278 Novel human DNA-binding protein #1

1                 (R,K){20,20}
                    (K){20}
         57: RGCSY KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         58: GCSYK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         59: CSYKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         60: SYKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         61: YKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         70: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
     71: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
     72: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXG

                (K){20}
     73: KKKKK KKKKKKKKKKKKKKKKKKKK KKXGX

                (K){20}
     74: KKKKK KKKKKKKKKKKKKKKKKKKK KXGX

                (K){20}
     75: KKKKK KKKKKKKKKKKKKKKKKKKK XGX

     AAU18279  ck: 3164  len: 181   ! Aau18279 Novel human DNA-binding protein #1

1             (R,K){20,20}
                (K){20}
    154: TRKPE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    155: RKPEK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    156: KPEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    157: PEKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
    158: EKKKK KKKKKKKKKKKKKKKKKKKK KKKX

                (K){20}
    159: KKKKK KKKKKKKKKKKKKKKKKKKK KKX

                (K){20}
    160: KKKKK KKKKKKKKKKKKKKKKKKKK KX

                (K){20}
    161: KKKKK KKKKKKKKKKKKKKKKKKKK X

     AAU18280  ck: 7117  len: 55    ! Aau18280 Novel human DNA-binding protein #1

1             (R,K){20,20}
                (K){20}
     22: DDKKN KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     23: DKKNK KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
     24: KKNKK KKKKKKKKKKKKKKKKKKKK KKKXG

                (K){20}
     25: KNKKK KKKKKKKKKKKKKKKKKKKK KKXGG

                (K){20}
     26: NKKKK KKKKKKKKKKKKKKKKKKKK KXGGR

                (K){20}
     27: KKKKK KKKKKKKKKKKKKKKKKKKK XGGRF

     AAU18281  ck: 9316  len: 67    ! Aau18281 Novel human DNA-binding protein #1

1             (R,K){20,20}
               (R,K){20}
     36: FTLQT RKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
     37: TLQTR KKKKKKKKKKKKKKKKKKKK KKKKX

                (K){20}
     38: LQTRK KKKKKKKKKKKKKKKKKKKK KKKXG

                (K){20}
     39: QTRKK KKKKKKKKKKKKKKKKKKKK KKXGG

                (K){20}
     40: TRKKK KKKKKKKKKKKKKKKKKKKK KXGGR

                (K){20}
     41: RKKKK KKKKKKKKKKKKKKKKKKKK XGGRF

     AAM53862  ck: 5383  len: 86    ! Aam53862 Human brain expressed single exon

1             (R,K){20,20}
               (R,K){20}
     15: RRRRG RRRRRKKKKKRKKKKKRKK KKRRR

               (R,K){20}
     16: RRRGR RRRRKKKKKRKKKKKRKKK KRRRR

               (R,K){20}
     17: RRGRR RRRKKKKKRKKKKKRKKKK RRRRR

               (R,K){20}
     18: RGRRR RRKKKKKRKKKKKRKKKKR RRRRR

               (R,K){20}
     19: GRRRR RKKKKKRKKKKKRKKKKRR RRRRR

               (R,K){20}
     20: RRRRR KKKKKRKKKKKRKKKKRRR RRRRR

               (R,K){20}
     21: RRRRK KKKKRKKKKKRKKKKRRRR RRRRR

               (R,K){20}
     22: RRRKK KKKRKKKKKRKKKKRRRRR RRRRR

               (R,K){20}
     23: RRKKK KKRKKKKKRKKKKRRRRRR RRRRR

               (R,K){20}
     24: RKKKK KRKKKKKRKKKKRRRRRRR RRRRR

               (R,K){20}
     25: KKKKK RKKKKKRKKKKRRRRRRRR RRRRR

               (R,K){20}
     26: KKKKK RKKKKKRKKKKRRRRRRRR RRRRR

               (R,K){20}
     27: KKKKR KKKKRKKKKRRRRRRRRRR RRRRR

               (R,K){20}
     28: KKKRK KKKRKKKKRRRRRRRRRRR RRRRR

               (R,K){20}
     29: KKRKK KKRKKKKRRRRRRRRRRRR RRRRR

               (R,K){20}
     30: KRKKK KRKKKKRRRRRRRRRRRRR RRRRR

               (R,K){20}
     31: RKKKK RKKKKRRRRRRRRRRRRRR RRRRR

               (R,K){20}
     32: KKKKK RKKKKRRRRRRRRRRRRRR RRRRR

               (R,K){20}
```

```
     33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
     34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
     35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

                (R,K){20}
     36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
     41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                 (R){20}
     42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

                 (R){20}
     43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

                 (R){20}
     44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

                 (R){20}
     45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

     AAM54700   ck: 1334   len: 86    ! Aam54700 Human brain expressed single exon

1              (R,K){20,20}
                (R,K){20}
     57: EEEEG RRRKKKKRKRKKKKKKKKKK KKKKK

                (R,K){20}
     58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

                (R,K){20}
     59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

                (R,K){20}
     60: EGRRR KKKKRKRKKKKKKKKKKKKK KKKKK

                (R,K){20}
     61: GRRRK KKKRKRKKKKKKKKKKKKKK KKKKK

                (R,K){20}
     62: RRRKK KKRKRKKKKKKKKKKKKKKK KKKKK

                (R,K){20}
     63: RRKKK KRKRKKKKKKKKKKKKKKKK KKKK

                (R,K){20}
     64: RKKKK RKRKKKKKKKKKKKKKKKKK KKK

                (R,K){20}
     65: KKKKR KRKKKKKKKKKKKKKKKKKK KK

                (R,K){20}
     66: KKKRK RKKKKKKKKKKKKKKKKKKK K

                 (K){20}
     67: KKRKR KKKKKKKKKKKKKKKKKKKK

     AAM54795   ck: 9082   len: 167   ! Aam54795 Human brain expressed single exon

1              (R,K){20,20}
                (R,K){20}
     33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

                (R,K){20}
     34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

                (R,K){20}
     35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR

     AAM55320   ck: 2276   len: 89    ! Aam55320 Human brain expressed single exon

1              (R,K){20,20}
                (R,K){20}
     23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

                 (K){20}
     46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

                 (K){20}
     47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

                (R,K){20}
     48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

                (R,K){20}
     49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

     AAM55623   ck: 1939   len: 130   ! Aam55623 Human brain expressed single exon

1              (R,K){20,20}
                (R,K){20}
     42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR

                (R,K){20}
     43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

                (R,K){20}
     44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK

                (R,K){20}
     45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

                (R,K){20}
     46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

                (R,K){20}
     47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

                (R,K){20}
     48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

                (R,K){20}
     49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

                (R,K){20}
     50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

                (R,K){20}
     51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

                (R,K){20}
     52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

                (R,K){20}
     53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE
```

```
            (R,K){20}
54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE

AAM55931  ck: 3607  len: 88    ! Aam55931 Human brain expressed single exon

1           (R,K){20,20}
              (K){20}
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK

AAM56490  ck: 3937  len: 85    ! Aam56490 Human brain expressed single exon p

1           (R,K){20,20}
              (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

              (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

              (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

              (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

              (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

AAM58224  ck: 1560  len: 88    ! Aam58224 Human brain expressed single exon p

1           (R,K){20,20}
              (R){20}
43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

              (R){20}
```

```
44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

            (R){20}
58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

            (R){20}
59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

            (R){20}
60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

            (R){20}
61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

            (R){20}
62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

AAM58395  ck: 2324  len: 36    ! Aam58395 Human brain expressed single exon

1          (R,K){20,20}
            (R,K){20}
16: KERKT KKRKRKRRRRRKKRRRKKRR R

            (R,K){20}
17: ERKTK KRKRKRRRRRKKRRRKKRRR

AAM61063  ck: 8343  len: 66    ! Aam61063 Human brain expressed single exon

1          (R,K){20,20}
            (R,K){20}
 6: ETERE KKKKRKKKKKKKKKKKKKKK KKKNK

            (R,K){20}
 7: TEREK KKKRKKKKKKKKKKKKKKKK KKNKK

            (R,K){20}
 8: EREKK KKRKKKKKKKKKKKKKKKKK KNKKK

            (R,K){20}
 9: REKKK KRKKKKKKKKKKKKKKKKKK NKKKK

AAM63533  ck: 2394  len: 57    ! Aam63533 Human brain expressed single exon

1          (R,K){20,20}
             (R){20}
20: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
21: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRG

             (R){20}
22: EEGRR RRRRRRRRRRRRRRRRRRRR RRRGR

             (R){20}
23: EGRRR RRRRRRRRRRRRRRRRRRRR RRGRR

             (R){20}
24: GRRRR RRRRRRRRRRRRRRRRRRRR RGRRR

             (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR

AAM64090  ck: 4228  len: 24    ! Aam64090 Human brain expressed single exon

1          (R,K){20,20}
            (R,K){20}
 1:       RRRRRRRKKKRRRRRRRRRR RKRT

            (R,K){20}
 2:     R RRRRRRKKKRRRRRRRRRRR KRT

            (R,K){20}
 3:    RR RRRRRKKKRRRRRRRRRRRK RT

            (R,K){20}
 4:   RRR RRRRKKKRRRRRRRRRRRKR T

AAM64863  ck: 3301  len: 52    ! Aam64863 Human brain expressed single exon

1          (R,K){20,20}
            (R,K){20}
12: KKKKN KKKKKKKKKRKRKRKRKKKK KKKKK

            (R,K){20}
13: KKKNK KKKKKKKKRKRKRKRKKKKK KKKKK

            (R,K){20}
14: KKNKK KKKKKKKRKRKRKRKKKKKK KKKKK

            (R,K){20}
15: KNKKK KKKKKKRKRKRKRKKKKKKK KKKKK

            (R,K){20}
16: NKKKK KKKKKRKRKRKRKKKKKKKK KKKKK

            (R,K){20}
17: KKKKK KKKKRKRKRKRKKKKKKKKK KKKKK

            (R,K){20}
18: KKKKK KKKRKRKRKRKKKKKKKKKK KKKKK

            (R,K){20}
```

```
19: KKKKK KKRKRKRKRKKKKKKKKKK KKKKK

          (R,K){20}
20: KKKKK KRKRKRKRKKKKKKKKKKK KKKKK

          (R,K){20}
21: KKKKK RKRKRKRKKKKKKKKKKKK KKKKK

          (R,K){20}
22: KKKKR KRKRKRKKKKKKKKKKKKK KKKKK

          (R,K){20}
23: KKKRK RKRKRKKKKKKKKKKKKKK KKKKK

          (R,K){20}
24: KKRKR KRKRKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
25: KRKRK RKRKKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
26: RKRKR KRKKKKKKKKKKKKKKKKK KKKKK

          (R,K){20}
27: KRKRK RKKKKKKKKKKKKKKKKKK KKKKA

           (K){20}
28: RKRKR KKKKKKKKKKKKKKKKKKK KKKAF

           (K){20}
29: KRKRK KKKKKKKKKKKKKKKKKKK KKAF

           (K){20}
30: RKRKK KKKKKKKKKKKKKKKKKKK KAF

           (K){20}
31: KRKKK KKKKKKKKKKKKKKKKKKK AF

AAM65355   ck: 4895   len: 51    ! Aam65355 Human brain expressed single exon

1          (R,K){20,20}
          (R,K){20}
18: LFKPM RKRRRKKKRRRKKRKKKRRR KKLTT

          (R,K){20}
19: FKPMR KRRRKKKRRRKKRKKKRRRK KLTTT

          (R,K){20}
20: KPMRK RRRKKKRRRKKRKKKRRRKK LTTTT

AAM66249   ck: 5383   len: 86    ! Aam66249 Human bone marrow expressed probe

1          (R,K){20,20}
          (R,K){20}
15: RRRRG RRRRRKKKKKKRKKKKKRKK KKRRR

          (R,K){20}
16: RRRGR RRRRKKKKKKRKKKKKRKKK KRRRR

          (R,K){20}
17: RRGRR RRRKKKKKKRKKKKKRKKKK RRRRR

          (R,K){20}
18: RGRRR RRKKKKKKRKKKKKRKKKKR RRRRR

          (R,K){20}
19: GRRRR RKKKKKKRKKKKKRKKKKRR RRRRR

          (R,K){20}
20: RRRRR KKKKKKRKKKKKRKKKKRRR RRRRR

          (R,K){20}
21: RRRRK KKKKKRKKKKKRKKKKRRRR RRRRR

          (R,K){20}
22: RRRKK KKKKRKKKKKRKKKKRRRRR RRRRR

          (R,K){20}
23: RRKKK KKKRKKKKKRKKKKRRRRRR RRRRR

          (R,K){20}
24: RKKKK KKRKKKKKRKKKKRRRRRRR RRRRR

          (R,K){20}
25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

          (R,K){20}
26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

          (R,K){20}
27: KKKKR KKKKKRKKKKRRRRRRRRRR RRRRR

          (R,K){20}
28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

          (R,K){20}
29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

          (R,K){20}
30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

          (R,K){20}
31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

          (R,K){20}
32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

          (R,K){20}
33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

          (R,K){20}
34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

          (R,K){20}
35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

          (R,K){20}
36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

           (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

           (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

           (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

           (R){20}
```

```
     45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

     AAM67100  ck: 1334  len: 86    ! Aam67100 Human bone marrow expressed probe

1              (R,K){20,20}
                (R,K){20}
     57: EEEEG RRRKKKKRKRKKKKKKKKKK KKKKK

                (R,K){20}
     58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

                (R,K){20}
     59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

                (R,K){20}
     60: EGRRR KKKKRKRKKKKKKKKKKKK KKKKK

                (R,K){20}
     61: GRRRK KKKRKRKKKKKKKKKKKKK KKKKK

                (R,K){20}
     62: RRRKK KKRKRKKKKKKKKKKKKKK KKKKK

                (R,K){20}
     63: RRKKK KRKRKKKKKKKKKKKKKKK KKKK

                (R,K){20}
     64: RKKKK RKRKKKKKKKKKKKKKKKK KKK

                (R,K){20}
     65: KKKKR KRKKKKKKKKKKKKKKKKK KK

                (R,K){20}
     66: KKKRK RKKKKKKKKKKKKKKKKKK K

                 (K){20}
     67: KKRKR KKKKKKKKKKKKKKKKKKKK

     AAM67180  ck: 9082  len: 167   ! Aam67180 Human bone marrow expressed probe

1              (R,K){20,20}
                (R,K){20}
     33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

                (R,K){20}
     34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

                (R,K){20}
     35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR

     AAM67717  ck: 2276  len: 89   ! Aam67717 Human bone marrow expressed probe

1              (R,K){20,20}
                (R,K){20}
     23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

                 (K){20}
     46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

                 (K){20}
     47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

                (R,K){20}
     48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

                (R,K){20}
     49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

     AAM68007  ck: 1939  len: 130   ! Aam68007 Human bone marrow expressed probe

1              (R,K){20,20}
                (R,K){20}
     42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR

                (R,K){20}
     43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

                (R,K){20}
     44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK

                (R,K){20}
     45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

                (R,K){20}
     46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

                (R,K){20}
     47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

                (R,K){20}
     48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

                (R,K){20}
     49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

                (R,K){20}
     50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

                (R,K){20}
     51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

                (R,K){20}
     52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

                (R,K){20}
     53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE

                (R,K){20}
     54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE

     AAM68298  ck: 3607  len: 88   ! Aam68298 Human bone marrow expressed probe

1              (R,K){20,20}
                 (K){20}
     39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
     47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
```

```
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

            (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

            (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

            (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK K

            (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK

AAM68869  ck: 3937  len: 85    ! Aam68869 Human bone marrow expressed probe

1         (R,K){20,20}
            (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
```

```
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

            (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

            (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

            (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

            (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

AAM69569  ck: 2686  len: 71    ! Aam69569 Human bone marrow expressed probe

1          (R,K){20,20}
            (K){20}
20: KKKKQ KKKKKKKKKKKKKKKKKKKK KKKKR

            (K){20}
21: KKKQK KKKKKKKKKKKKKKKKKKKK KKKRK

            (K){20}
22: KKQKK KKKKKKKKKKKKKKKKKKKK KKRKR

            (K){20}
23: KQKKK KKKKKKKKKKKKKKKKKKKK KRKRK

            (K){20}
24: QKKKK KKKKKKKKKKKKKKKKKKKK RKRKK

           (R,K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKR KRKKK

           (R,K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKRK RKKKK

           (R,K){20}
27: KKKKK KKKKKKKKKKKKKKKKKRKR KKKKK

           (R,K){20}
28: KKKKK KKKKKKKKKKKKKKKKRKRK KKKKK

           (R,K){20}
29: KKKKK KKKKKKKKKKKKKKKRKRKK KKKKK

           (R,K){20}
30: KKKKK KKKKKKKKKKKKKKRKRKKK KKKKK

           (R,K){20}
31: KKKKK KKKKKKKKKKKKKRKRKKKK KKKKK

           (R,K){20}
32: KKKKK KKKKKKKKKKKKRKRKKKKK KKKKK

           (R,K){20}
```

```
         33: KKKKK KKKKKKKKKKKRKRKKKKKK KKKKK

                     (R,K){20}
         34: KKKKK KKKKKKKKKKRKRKKKKKKK KKKKK

                     (R,K){20}
         35: KKKKK KKKKKKKKKRKRKKKKKKKK KKKKK

                     (R,K){20}
         36: KKKKK KKKKKKKKRKRKKKKKKKKK KKKKK

                     (R,K){20}
         37: KKKKK KKKKKKKRKRKKKKKKKKKK KKKKK

                     (R,K){20}
         38: KKKKK KKKKKKRKRKKKKKKKKKKK KKKKK

                     (R,K){20}
         39: KKKKK KKKKKRKRKKKKKKKKKKKK KKKKK

                     (R,K){20}
         40: KKKKK KKKKRKRKKKKKKKKKKKKK KKKKK

                     (R,K){20}
         41: KKKKK KKKRKRKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
         42: KKKKK KKRKRKKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
         43: KKKKK KRKRKKKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
         44: KKKKK RKRKKKKKKKKKKKKKKKKK KKKKK

                     (R,K){20}
         45: KKKKR KRKKKKKKKKKKKKKKKKKK KKKKS

                     (R,K){20}
         46: KKKRK RKKKKKKKKKKKKKKKKKKK KKKSA

                      (K){20}
         47: KKRKR KKKKKKKKKKKKKKKKKKKK KKSAH

                      (K){20}
         48: KRKRK KKKKKKKKKKKKKKKKKKKK KSAH

                      (K){20}
         49: RKRKK KKKKKKKKKKKKKKKKKKKK SAH

         AAM70678  ck: 1560  len: 88    ! Aam70678 Human bone marrow expressed probe

1                  (R,K){20,20}
                      (R){20}
         43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
```

```
         49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                      (R){20}
         58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                      (R){20}
         59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

                      (R){20}
         60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

                      (R){20}
         61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

                      (R){20}
         62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

         AAM70881  ck: 2324  len: 36    ! Aam70881 Human bone marrow expressed probe

1                  (R,K){20,20}
                    (R,K){20}
         16: KERKT KKRKRKRRRRRKKRRRKKRR R

                    (R,K){20}
         17: ERKTK KRKRKRRRRRKKRRRKKRRR

         AAM73767  ck: 8343  len: 66    ! Aam73767 Human bone marrow expressed probe

1                  (R,K){20,20}
                    (R,K){20}
          6: ETERE KKKKRKKKKKKKKKKKKKKK KKKNK

                    (R,K){20}
          7: TEREK KKKRKKKKKKKKKKKKKKKK KKNKK

                    (R,K){20}
          8: EREKK KKRKKKKKKKKKKKKKKKKK KNKKK

                    (R,K){20}
          9: REKKK KRKKKKKKKKKKKKKKKKKK NKKKK

         AAM76347  ck: 2394  len: 57    ! Aam76347 Human bone marrow expressed probe

1                  (R,K){20,20}
                      (R){20}
         20: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRR
```

```
                    (R){20}
       21: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRG

                    (R){20}
       22: EEGRR RRRRRRRRRRRRRRRRRRRR RRRGR

                    (R){20}
       23: EGRRR RRRRRRRRRRRRRRRRRRRR RRGRR

                    (R){20}
       24: GRRRR RRRRRRRRRRRRRRRRRRRR RGRRR

                    (R){20}
       25: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR

       AAM76911  ck: 4228  len: 24    ! Aam76911 Human bone marrow expressed probe

1                 (R,K){20,20}
                   (R,K){20}
        1:       RRRRRRRKKKRRRRRRRRRR RKRT

                   (R,K){20}
        2:     R RRRRRRKKKRRRRRRRRRRR KRT

                   (R,K){20}
        3:    RR RRRRRKKKRRRRRRRRRRRK RT

                   (R,K){20}
        4:   RRR RRRRKKKRRRRRRRRRRRKR T

       AAM78048  ck: 4895  len: 51    ! Aam78048 Human bone marrow expressed probe

1                 (R,K){20,20}
                   (R,K){20}
       18: LFKPM RKRRRKKKRRRKKRKKKRRR KKLTT

                   (R,K){20}
       19: FKPMR KRRRKKKRRRKKRKKKRRRK KLTTT

                   (R,K){20}
       20: KPMRK RRRKKKRRRKKRKKKRRRKK LTTTT

       AAM82533  ck: 1736  len: 40    ! Aam82533 Human immune/haematopoietic antige

1                 (R,K){20,20}
                    (K){20}
       18: LPGSL KKKKKKKKKKKKKKKKKKKK KKK

                    (K){20}
       19: PGSLK KKKKKKKKKKKKKKKKKKKK KK

                    (K){20}
       20: GSLKK KKKKKKKKKKKKKKKKKKKK K

                    (K){20}
       21: SLKKK KKKKKKKKKKKKKKKKKKKK

       AAM85748  ck: 7503  len: 74    ! Aam85748 Human immune/haematopoietic antige

1                 (R,K){20,20}
                    (K){20}
       45: SHLTD KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       46: HLTDK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       47: LTDKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       48: TDKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       49: DKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                    (K){20}
       52: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                    (K){20}
       53: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                    (K){20}
       54: KKKKK KKKKKKKKKKKKKKKKKKKK K

                    (K){20}
       55: KKKKK KKKKKKKKKKKKKKKKKKKK

       AAM90546  ck: 6676  len: 74    ! Aam90546 Human immune/haematopoietic antiger

1                 (R,K){20,20}
                    (K){20}
       47: LRTFQ KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       48: RTFQK KKKKKKKKKKKKKKKKKKKK KKKKX

                    (K){20}
       49: TFQKK KKKKKKKKKKKKKKKKKKKK KKKXX

                    (K){20}
       50: FQKKK KKKKKKKKKKKKKKKKKKKK KKXXG

                    (K){20}
       51: QKKKK KKKKKKKKKKKKKKKKKKKK KXXG

                    (K){20}
       52: KKKKK KKKKKKKKKKKKKKKKKKKK XXG

       AAM90618  ck: 5691  len: 108   ! Aam90618 Human immune/haematopoietic antiger

1                 (R,K){20,20}
                    (K){20}
       78: VRPCL KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       79: RPCLK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       80: PCLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       81: CLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
       84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
```

```
              (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKKK

AAM91030   ck: 8102   len: 111   ! Aam91030 Human immune/haematopoietic antige

1          (R,K){20,20}
              (K){20}
78: EFHIL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
79: FHILK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
80: HILKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
81: ILKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
87: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
90: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKK

AAM91162   ck: 1109   len: 98   ! Aam91162 Human immune/haematopoietic antige

1          (R,K){20,20}
              (K){20}
53: QTKNT KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
54: TKNTK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
55: KNTKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
56: NTKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
57: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGR

              (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KGGRS

              (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK GGRSR

AAM91891   ck: 8102   len: 111   ! Aam91891 Human immune/haematopoietic antige

1          (R,K){20,20}
              (K){20}
78: EFHIL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
79: FHILK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
80: HILKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
81: ILKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
82: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
86: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                 (K){20}
       87: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                 (K){20}
       89: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                 (K){20}
       90: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                 (K){20}
       91: KKKKK KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
       92: KKKKK KKKKKKKKKKKKKKKKKKKK

       AAM91933   ck: 8285   len: 55    ! Aam91933 Human digestive system antigen SEQ

1               (R,K){20,20}
                 (K){20}
       33: PPTRP KKKKKKKKKKKKKKKKKKKK KGK

                 (K){20}
       34: PTRPK KKKKKKKKKKKKKKKKKKKK GK

       AAM92433   ck: 1663   len: 87    ! Aam92433 Human digestive system antigen SEQ

1               (R,K){20,20}
                 (K){20}
       36: KWSSX KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       37: WSSXK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       38: SSXKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       39: SXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       40: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                 (K){20}
       50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                 (K){20}
       59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                 (K){20}
       60: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGX

                 (K){20}
       61: KKKKK KKKKKKKKKKKKKKKKKKKK KRGXP

                 (K){20}
       62: KKKKK KKKKKKKKKKKKKKKKKKKK RGXPF

                (R,K){20}
       63: KKKKK KKKKKKKKKKKKKKKKKKKR GXPFX

       AAO00092   ck: 9065   len: 113   ! Aao00092 Human polypeptide SEQ ID NO 13984.

1               (R,K){20,20}
                 (K){20}
       91: CLGCL KKKKKKKKKKKKKKKKKKKK KFF

                 (K){20}
       92: LGCLK KKKKKKKKKKKKKKKKKKKK FF

       AAO00222   ck: 2916   len: 132   ! Aao00222 Human polypeptide SEQ ID NO 14114.

1               (R,K){20,20}
                 (K){20}
       28: XPLPP KKKKKKKKKKKKKKKKKKKK GGPPK

       AAO00232   ck: 1000   len: 102   ! Aao00232 Human polypeptide SEQ ID NO 14124.

1               (R,K){20,20}
                 (K){20}
       29: RDCFF KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       30: DCFFK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       31: CFFKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
32: FFKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
33: FKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

              (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KGGPF

              (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK GGPFK
```

```
AAO00291   ck: 8100   len: 124   ! Aao00291 Human polypeptide SEQ ID NO 14183.

 1             (R,K){20,20}
                (K){20}
22: QHFCM KKKKKKKKKKKKKKKKKKKK FFKKG

AAO00439   ck: 6396   len: 122   ! Aao00439 Human polypeptide SEQ ID NO 14331.

 1             (R,K){20,20}
                (K){20}
23: CLWLV KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
24: LWLVK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
25: WLVKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
26: LVKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
27: VKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRK

                (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKRKN

                (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KRKNY

                (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK RKNYF

              (R,K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKR KNYFH
```

```
                  (R,K){20}
        44: KKKKK KKKKKKKKKKKKKKKKKKRK NYFHV

        AA000608  ck: 7114  len: 57    ! Aao00608 Human polypeptide SEQ ID NO 14500.
1                 (R,K){20,20}
                    (K){20}
        22: SLSPE KKKKKKKKKKKKKKKKKKKK KSASS

                    (K){20}
        23: LSPEK KKKKKKKKKKKKKKKKKKKK SASSS

        AA000619  ck: 6838  len: 117   ! Aao00619 Human polypeptide SEQ ID NO 14511.
1                 (R,K){20,20}
                    (K){20}
         1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
         5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKD

                    (K){20}
         6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKDG

                    (K){20}
         7: KKKKK KKKKKKKKKKKKKKKKKKKK KKDGG

                    (K){20}
         8: KKKKK KKKKKKKKKKKKKKKKKKKK KDGGG

                    (K){20}
         9: KKKKK KKKKKKKKKKKKKKKKKKKK DGGGA

        AA001368  ck: 3955  len: 123   ! Aao01368 Human polypeptide SEQ ID NO 15260.
1                 (R,K){20,20}
                    (K){20}
        87: RQLIT KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        88: QLITK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        89: LITKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        90: ITKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        91: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        92: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKD

                    (K){20}
        95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKDD

                    (K){20}
        96: KKKKK KKKKKKKKKKKKKKKKKKKK KKDDS

                    (K){20}
        97: KKKKK KKKKKKKKKKKKKKKKKKKK KDDSG

                    (K){20}
        98: KKKKK KKKKKKKKKKKKKKKKKKKK DDSGK

        AA001560  ck: 7002  len: 74    ! Aao01560 Human polypeptide SEQ ID NO 15452.
1                 (R,K){20,20}
                    (K){20}
        30: KLYHL KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        31: LYHLK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        32: YHLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        33: HLKKK KKKKKKKKKKKKKKKKKKKK KKKKT

                    (K){20}
        34: LKKKK KKKKKKKKKKKKKKKKKKKK KKKTK

                    (K){20}
        35: KKKKK KKKKKKKKKKKKKKKKKKKK KKTKK

                    (K){20}
        36: KKKKK KKKKKKKKKKKKKKKKKKKK KTKKK

                    (K){20}
        37: KKKKK KKKKKKKKKKKKKKKKKKKK TKKKN

        AA002135  ck: 8265  len: 74    ! Aao02135 Human polypeptide SEQ ID NO 16027.
1                 (R,K){20,20}
                    (K){20}
        41: MPPPP KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        42: PPPPK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        43: PPPKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        44: PPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        45: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                    (K){20}
        50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKQ
```

```
          (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKQ

          (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKQ

          (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KQ

          (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK Q

AAO02186  ck: 4844  len: 57   ! Aao02186 Human polypeptide SEQ ID NO 16078.

1         (R,K){20,20}
          (K){20}
11: HCCLL KKKKKKKKKKKKKKKKKKKK KIKNK

          (K){20}
12: CCLLK KKKKKKKKKKKKKKKKKKKK IKNKK

AAO02310  ck: 704  len: 137   ! Aao02310 Human polypeptide SEQ ID NO 16202.

1         (R,K){20,20}
          (R,K){20}
22: HSLNL KKKKKKKKKKKKKKKKKKKR GGGVK

AAO02477  ck: 2018  len: 112   ! Aao02477 Human polypeptide SEQ ID NO 16369.

1         (R,K){20,20}
          (K){20}
83: LASAV KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
84: ASAVK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
85: SAVKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
86: AVKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
87: VKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
88: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
89: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

          (K){20}
90: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

          (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKKK KK

          (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKK K

          (K){20}
93: KKKKK KKKKKKKKKKKKKKKKKKKK

AAO02733  ck: 2860  len: 132   ! Aao02733 Human polypeptide SEQ ID NO 16625.

1         (R,K){20,20}
          (K){20}
86: FFFSL KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
87: FFSLK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
88: FSLKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
89: SLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
90: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
91: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
92: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
93: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
94: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
95: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

          (K){20}
97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRA

          (K){20}
98: KKKKK KKKKKKKKKKKKKKKKKKKK KKRAR

          (K){20}
99: KKKKK KKKKKKKKKKKKKKKKKKKK KRARS

           (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK RARSR

          (R,K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKR ARSRS

AAO02946  ck: 3649  len: 126   ! Aao02946 Human polypeptide SEQ ID NO 16838.

1         (R,K){20,20}
          (K){20}
25: DEATS KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
26: EATSK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
27: ATSKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
28: TSKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
29: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
   33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
   35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
   36: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                (K){20}
   37: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGF

                (K){20}
   38: KKKKK KKKKKKKKKKKKKKKKKKKK GGGFK

   AAO02961  ck: 4320  len: 83   ! Aao02961 Human polypeptide SEQ ID NO 16853.

1             (R,K){20,20}
                (K){20}
   22: RMFSS KKKKKKKKKKKKKKKKKKKK KTAIT

                (K){20}
   23: MFSSK KKKKKKKKKKKKKKKKKKKK TAITK

   AAO03006  ck: 5212  len: 102   ! Aao03006 Human polypeptide SEQ ID NO 16898.

1             (R,K){20,20}
                (K){20}
   59: HDFFP KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   60: DFFPK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   61: FFPKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   62: FPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   63: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKA

                (K){20}
   64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKAS

                (K){20}
   65: KKKKK KKKKKKKKKKKKKKKKKKKK KKASS

                (K){20}
   66: KKKKK KKKKKKKKKKKKKKKKKKKK KASSS

                (K){20}
   67: KKKKK KKKKKKKKKKKKKKKKKKKK ASSSS

   AAO03024  ck: 6865  len: 64   ! Aao03024 Human polypeptide SEQ ID NO 16916.

1             (R,K){20,20}
               (R,K){20}
   15: SFAKA RKKKKKKKKKKKKKKKKKKK KRGGP

                (K){20}
   16: FAKAR KKKKKKKKKKKKKKKKKKKK RGGPP

               (R,K){20}
   17: AKARK KKKKKKKKKKKKKKKKKKKR GGPPL

   AAO03113  ck: 5587  len: 60   ! Aao03113 Human polypeptide SEQ ID NO 17005.

1             (R,K){20,20}
                (K){20}
   33: EINML KKKKKKKKKKKKKKKKKKKK KKKRG

                (K){20}
   34: INMLK KKKKKKKKKKKKKKKKKKKK KKRGG

                (K){20}
   35: NMLKK KKKKKKKKKKKKKKKKKKKK KRGGG

                (K){20}
   36: MLKKK KKKKKKKKKKKKKKKKKKKK RGGGL

               (R,K){20}
   37: LKKKK KKKKKKKKKKKKKKKKKKKR GGGL

   AAO03123  ck: 6627  len: 108   ! Aao03123 Human polypeptide SEQ ID NO 17015.

1             (R,K){20,20}
                (K){20}
   33: RTWRX KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   34: TWRXK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   35: WRXKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   36: RXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   37: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                (K){20}
   39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                (K){20}
   40: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGG

                (K){20}
   41: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGG

                (K){20}
   42: KKKKK KKKKKKKKKKKKKKKKKKKK RGGGQ

               (R,K){20}
   43: KKKKK KKKKKKKKKKKKKKKKKKKR GGGQK

   AAO03132  ck: 3903  len: 116   ! Aao03132 Human polypeptide SEQ ID NO 17024.

1             (R,K){20,20}
                (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKR

                (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKRG

                (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKRGG

                (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KRGGP

                (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK RGGPP
```

```
              (R,K){20}
  6: KKKKK KKKKKKKKKKKKKKKKKKKR GGPPK

AAO03152   ck: 7891   len: 35    ! Aao03152 Human polypeptide SEQ ID NO 17044.

1           (R,K){20,20}
              (R,K){20}
 13: LIXYT RKKKKKKKKKKKKKKKKKKK RGG

              (R,K){20}
 14: IXYTR KKKKKKKKKKKKKKKKKKKR GG

AAO03168   ck: 7233   len: 67    ! Aao03168 Human polypeptide SEQ ID NO 17060.

1           (R,K){20,20}
                (K){20}
 17: ELLMP KKKKKKKKKKKKKKKKKKKK KRGGA

                (K){20}
 18: LLMPK KKKKKKKKKKKKKKKKKKKK RGGAP

              (R,K){20}
 19: LMPKK KKKKKKKKKKKKKKKKKKKR GGAPF

AAO03243   ck: 411    len: 95    ! Aao03243 Human polypeptide SEQ ID NO 17135.

1           (R,K){20,20}
                (K){20}
 51: LLHAV KKKKKKKKKKKKKKKKKKKK KKEKK

                (K){20}
 52: LHAVK KKKKKKKKKKKKKKKKKKKK KEKKR

                (K){20}
 53: HAVKK KKKKKKKKKKKKKKKKKKKK EKKRG

AAO03273   ck: 2981   len: 88    ! Aao03273 Human polypeptide SEQ ID NO 17165.

1           (R,K){20,20}
                (K){20}
 43: IXYLL KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 44: XYLLK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 45: YLLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 46: LLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 47: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKP

                (K){20}
 54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKPT

                (K){20}
 55: KKKKK KKKKKKKKKKKKKKKKKKKK KKPTP

                (K){20}
 56: KKKKK KKKKKKKKKKKKKKKKKKKK KPTPK

                (K){20}
 57: KKKKK KKKKKKKKKKKKKKKKKKKK PTPKK

AAO03277   ck: 2981   len: 89    ! Aao03277 Human polypeptide SEQ ID NO 17169.

1           (R,K){20,20}
                (K){20}
  1:       KKKKKKKKKKKKKKKKKKKK KKKRM

                (K){20}
  2:     K KKKKKKKKKKKKKKKKKKKK KKRMK

                (K){20}
  3:    KK KKKKKKKKKKKKKKKKKKKK KRMKN

                (K){20}
  4:   KKK KKKKKKKKKKKKKKKKKKKK RMKNK

              (R,K){20}
  5:  KKKK KKKKKKKKKKKKKKKKKKKR MKNKK

AAO03278   ck: 6777   len: 74    ! Aao03278 Human polypeptide SEQ ID NO 17170.

1           (R,K){20,20}
                (K){20}
 47: RAVAQ KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
 48: AVAQK KKKKKKKKKKKKKKKKKKKK KKGGG

                (K){20}
 49: VAQKK KKKKKKKKKKKKKKKKKKKK KGGGV

                (K){20}
 50: AQKKK KKKKKKKKKKKKKKKKKKKK GGGVL

AAO03284   ck: 9650   len: 115   ! Aao03284 Human polypeptide SEQ ID NO 17176.

1           (R,K){20,20}
                (K){20}
 22: KALVS KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 23: ALVSK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
 24: LVSKK KKKKKKKKKKKKKKKKKKKK KKKKR

                (K){20}
 25: VSKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                (K){20}
 26: SKKKK KKKKKKKKKKKKKKKKKKKK KKRGG

                (K){20}
 27: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGA
```

```
              (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK RGGAL

             (R,K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKR GGALK

AAO03288  ck: 1109  len: 117   ! Aao03288 Human polypeptide SEQ ID NO 17180.
1            (R,K){20,20}
              (K){20}
31: SCSYP KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
32: CSYPK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
33: SYPKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
34: YPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
35: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

              (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRK

              (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKRKR

              (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KRKRK

              (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK RKRKK

             (R,K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKR KRKKK

             (R,K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKRK RKKKK

             (R,K){20}
47: KKKKK KKKKKKKKKKKKKKKKKRKR KKKKK

             (R,K){20}
48: KKKKK KKKKKKKKKKKKKKKKRKRK KKKKK

             (R,K){20}
49: KKKKK KKKKKKKKKKKKKKKRKRKK KKKKT

             (R.K){20}
50: KKKKK KKKKKKKKKKKKKKRKRKKK KKKTX

             (R,K){20}
51: KKKKK KKKKKKKKKKKKKRKRKKKK KKTXG

             (R,K){20}
52: KKKKK KKKKKKKKKKKKRKRKKKKK KTXGG

             (R,K){20}
53: KKKKK KKKKKKKKKKKRKRKKKKKK TXGGG

AAO03375  ck: 693   len: 54    ! Aao03375 Human polypeptide SEQ ID NO 17267.
1            (R,K){20,20}
              (K){20}
25: TXLEE KKKKKKKKKKKKKKKKKKKK KKKAP

              (K){20}
26: XLEEK KKKKKKKKKKKKKKKKKKKK KKAPG

              (K){20}
27: LEEKK KKKKKKKKKKKKKKKKKKKK KAPGG

              (K){20}
28: EEKKK KKKKKKKKKKKKKKKKKKKK APGGG

AAO03459  ck: 5217  len: 58    ! Aao03459 Human polypeptide SEQ ID NO 17351.
1            (R,K){20,20}
              (K){20}
28: XMCVF KKKKKKKKKKKKKKKKKKKK KDPPK

              (K){20}
29: MCVFK KKKKKKKKKKKKKKKKKKKK DPPKK

AAO03475  ck: 315   len: 138   ! Aao03475 Human polypeptide SEQ ID NO 17367.
1            (R,K){20,20}
              (K){20}
44: LLXLT KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
45: LXLTK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
46: XLTKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
47: LTKKK KKKKKKKKKKKKKKKKKKKK KKGGP

              (K){20}
48: TKKKK KKKKKKKKKKKKKKKKKKKK KGGPR

              (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK GGPRG

AAO03615  ck: 7613  len: 128   ! Aao03615 Human polypeptide SEQ ID NO 17507.
1            (R,K){20,20}
              (K){20}
28: PLPRX KKKKKKKKKKKKKKKKKKKK KKTAA

              (K){20}
29: LPRXK KKKKKKKKKKKKKKKKKKKK KTAAG

              (K){20}
30: PRXKK KKKKKKKKKKKKKKKKKKKK TAAGG

AAO03623  ck: 3884  len: 116   ! Aao03623 Human polypeptide SEQ ID NO 17515.
1            (R,K){20,20}
              (K){20}
77: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                  (K){20}
       78: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       79: SRAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       80: RAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       81: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       82: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       83: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       84: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKP

                  (K){20}
       85: KKKKK KKKKKKKKKKKKKKKKKKKK KKKPG

                  (K){20}
       86: KKKKK KKKKKKKKKKKKKKKKKKKK KKPGG

                  (K){20}
       87: KKKKK KKKKKKKKKKKKKKKKKKKK KPGGG

                  (K){20}
       88: KKKKK KKKKKKKKKKKKKKKKKKKK PGGGV

       AA003700   ck: 1279  len: 43    ! Aa003700 Human polypeptide SEQ ID NO 17592.

1                (R,K){20,20}
                  (K){20}
       19: YSQRL KKKKKKKKKKKKKKKKKKKK GGGPF

       AA003703   ck: 7016  len: 113   ! Aa003703 Human polypeptide SEQ ID NO 17595.

1                (R,K){20,20}
                  (K){20}
       58: IWDAE KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       59: WDAEK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       60: DAEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       61: AEKKK KKKKKKKKKKKKKKKKKKKK KKKKS

                  (K){20}
       62: EKKKK KKKKKKKKKKKKKKKKKKKK KKKSP

                  (K){20}
       63: KKKKK KKKKKKKKKKKKKKKKKKKK KKSPG

                  (K){20}
       64: KKKKK KKKKKKKKKKKKKKKKKKKK KSPGG

                  (K){20}
       65: KKKKK KKKKKKKKKKKKKKKKKKKK SPGGA

       AA003766   ck: 8808  len: 81    ! Aa003766 Human polypeptide SEQ ID NO 17658.

1                (R,K){20,20}

                  (K){20}
       49: TTTAC KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       50: TTACK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       51: TACKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       52: ACKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       53: CKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                  (K){20}
       54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                  (K){20}
       55: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                  (K){20}
       56: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGG

                  (K){20}
       57: KKKKK KKKKKKKKKKKKKKKKKKKK GGGGA

       AA003841   ck: 8734  len: 100   ! Aa003841 Human polypeptide SEQ ID NO 17733.

1                (R,K){20,20}
                  (K){20}
       36: KQHYP KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       37: QHYPK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       38: HYPKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       39: YPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       40: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                  (K){20}
       41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRT

                  (K){20}
       42: KKKKK KKKKKKKKKKKKKKKKKKKK KKRTK

                  (K){20}
       43: KKKKK KKKKKKKKKKKKKKKKKKKK KRTKQ

                  (K){20}
       44: KKKKK KKKKKKKKKKKKKKKKKKKK RTKQK

                (R,K){20}
       45: KKKKK KKKKKKKKKKKKKKKKKKKR TKQKK

       AA003906   ck: 4312  len: 100   ! Aa003906 Human polypeptide SEQ ID NO 17798.

1                (R,K){20,20}
                  (K){20}
       34: NKQNQ KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       35: KQNQK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
       36: QNQKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
37: NQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
38: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKN

              (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKNG

              (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKNGG

              (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KNGGG

              (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK NGGGG

AAO03967  ck: 7917  len: 53    ! Aao03967 Human polypeptide SEQ ID NO 17859.
1
            (R,K){20,20}
              (K){20}
16: KLXKS KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
17: LXKSK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
18: XKSKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
19: KSKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
20: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKS

              (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKSR

              (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKSRG

              (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KSRGG

              (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK SRGGA

AAO04512  ck: 4277  len: 62    ! Aao04512 Human polypeptide SEQ ID NO 18404.
1
            (R,K){20,20}
              (K){20}
24: KKKKE KKKKKKKKKKKKKKKKKKKK GGGGF

AAO04619  ck: 8654  len: 39    ! Aao04619 Human polypeptide SEQ ID NO 18511.
1
            (R,K){20,20}
              (K){20}
13: LITFL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
14: ITFLK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
15: TFLKK KKKKKKKKKKKKKKKKKKKK KKKKN

              (K){20}
16: FLKKK KKKKKKKKKKKKKKKKKKKK KKKN

              (K){20}
17: LKKKK KKKKKKKKKKKKKKKKKKKK KKN

              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KN

              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK N

AAO04644  ck: 2038  len: 70    ! Aao04644 Human polypeptide SEQ ID NO 18536.
1
            (R,K){20,20}
             (R,K){20}
23: AKPFT KKKKKKKKKKKKKKKKKKKR GGAPL

AAO04645  ck: 7825  len: 118   ! Aao04645 Human polypeptide SEQ ID NO 18537.
1
            (R,K){20,20}
             (R,K){20}
46: IKSFL RKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
47: KSFLR KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
48: SFLRK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
49: FLRKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
50: LRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
51: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKP
```

```
              (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKPR

              (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKPRG

              (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KPRGG

              (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK PRGGG

AAO04647   ck: 4805   len: 58    ! Aao04647 Human polypeptide SEQ ID NO 18539.
1            (R,K){20,20}
              (K){20}
28: KPTRP KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
29: PTRPK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
30: TRPKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
31: RPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
32: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKS

              (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKS

              (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKS

              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KS

              (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK S

AAO04674   ck: 2036   len: 33    ! Aao04674 Human polypeptide SEQ ID NO 18566.
1            (R,K){20,20}
              (K){20}
 9: WCYIT KKKKKKKKKKKKKKKKKKKK KKARG

              (K){20}
10: CYITK KKKKKKKKKKKKKKKKKKKK KARG

              (K){20}
11: YITKK KKKKKKKKKKKKKKKKKKKK ARG

AAO04679   ck: 7190   len: 31    ! Aao04679 Human polypeptide SEQ ID NO 18571.
1            (R,K){20,20}
              (K){20}
 8: CWFTQ KKKKKKKKKKKKKKKKKKKK KDRG

              (K){20}
 9: WFTQK KKKKKKKKKKKKKKKKKKKK DRG
```

```
AAO04682   ck: 5605   len: 60    ! Aao04682 Human polypeptide SEQ ID NO 18574.
1            (R,K){20,20}
              (K){20}
 2:     L KKKKKKKKKKKKKKKKKKKK KKKQK

              (K){20}
 3:    LK KKKKKKKKKKKKKKKKKKKK KKQKD

              (K){20}
 4:   LKK KKKKKKKKKKKKKKKKKKKK KQKDL

              (K){20}
 5:  LKKK KKKKKKKKKKKKKKKKKKKK QKDLK

AAO04690   ck: 7157   len: 81    ! Aao04690 Human polypeptide SEQ ID NO 18582.
1            (R,K){20,20}
              (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKP

              (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKPK

              (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKPKK

              (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KPKKK

              (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK PKKKK

AAO04715   ck: 6984   len: 35    ! Aao04715 Human polypeptide SEQ ID NO 18607.
1            (R,K){20,20}
              (K){20}
 8: LGSKD KKKKKKKKKKKKKKKKKKKK KKAKK

              (K){20}
 9: GSKDK KKKKKKKKKKKKKKKKKKKK KAKKK

              (K){20}
10: SKDKK KKKKKKKKKKKKKKKKKKKK AKKKD

AAO04743   ck: 9412   len: 54    ! Aao04743 Human polypeptide SEQ ID NO 18635.
1            (R,K){20,20}
              (K){20}
21: KINKL KKKKKKKKKKKKKKKKKKKK KRAAA

              (K){20}
22: INKLK KKKKKKKKKKKKKKKKKKKK RAAAR

             (R,K){20}
23: NKLKK KKKKKKKKKKKKKKKKKKKR AAARD

AAO04747   ck: 8399   len: 39    ! Aao04747 Human polypeptide SEQ ID NO 18639.
```

```
1                (R,K){20,20}
                   (K){20}
    13: RTGFV KKKKKKKKKKKKKKKKKKKK KKRGG

                   (K){20}
    14: TGFVK KKKKKKKKKKKKKKKKKKKK KRGGG

                   (K){20}
    15: GFVKK KKKKKKKKKKKKKKKKKKKK RGGGF

                 (R,K){20}
    16: FVKKK KKKKKKKKKKKKKKKKKKKR GGGF

    AA004752   ck: 1021   len: 71    ! Aao04752 Human polypeptide SEQ ID NO 18644.

1                (R,K){20,20}
                   (K){20}
    19: QEQGL KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    20: EQGLK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    21: QGLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    22: GLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    23: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                   (K){20}
    24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                   (K){20}
    25: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                   (K){20}
    26: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGL

                   (K){20}
    27: KKKKK KKKKKKKKKKKKKKKKKKKK GGGLL

    AA004755   ck: 5521   len: 59    ! Aao04755 Human polypeptide SEQ ID NO 18647.

1                (R,K){20,20}
                   (K){20}
    31: INSLE KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    32: NSLEK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    33: SLEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    34: LEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    35: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKT

                   (K){20}
    36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKT

                   (K){20}
    37: KKKKK KKKKKKKKKKKKKKKKKKKK KKT

                   (K){20}
    38: KKKKK KKKKKKKKKKKKKKKKKKKK KT

                   (K){20}
    39: KKKKK KKKKKKKKKKKKKKKKKKKK T

    AA004756   ck: 6952   len: 26    ! Aao04756 Human polypeptide SEQ ID NO 18648.

1                (R,K){20,20}
                   (K){20}
     4:   FFY KKKKKKKKKKKKKKKKKKKK SSS

    AA004758   ck: 7836   len: 115   ! Aao04758 Human polypeptide SEQ ID NO 18650.

1                (R,K){20,20}
                   (K){20}
     7: PFYQL KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
     8: FYQLK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
     9: YQLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    10: QLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    11: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
    26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
         (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKP

         (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKPQ

         (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKPQQ

         (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KPQQK

         (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK PQQKG
```

```
AA004764  ck: 2035  len: 54   ! Aao04764 Human polypeptide SEQ ID NO 18656.

1                (R,K){20,20}
                   (K){20}
 9: GDSSL KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
10: DSSLK KKKKKKKKKKKKKKKKKKKK KKKKI

                   (K){20}
11: SSLKK KKKKKKKKKKKKKKKKKKKK KKKIW

                   (K){20}
12: SLKKK KKKKKKKKKKKKKKKKKKKK KKIWE

                   (K){20}
13: LKKKK KKKKKKKKKKKKKKKKKKKK KIWEF

                   (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK IWEFW

AA004802  ck: 1223  len: 105  ! Aao04802 Human polypeptide SEQ ID NO 18694.

1                (R,K){20,20}
                   (K){20}
34: FFSRQ KKKKKKKKKKKKKKKKKKKK GGGGL

AA004854  ck: 4121  len: 58   ! Aao04854 Human polypeptide SEQ ID NO 18746.

1                (R,K){20,20}
                   (K){20}
13: TPFRA KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
14: PFRAK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
15: FRAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
16: RAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
17: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGG

                (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGN

                (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK RGGNF

               (R,K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKR GGNFK

AA004856  ck: 4861  len: 135   ! Aao04856 Human polypeptide SEQ ID NO 18748.
1             (R,K){20,20}
                (K){20}
41: LTLTT KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
42: TLTTK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
43: LTTKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
44: TTKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
45: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGG

                (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK GGGGP

AA004872  ck: 3224  len: 93    ! Aao04872 Human polypeptide SEQ ID NO 18764.
1             (R,K){20,20}
                (K){20}
 8: FLYKL KKKKKKKKKKKKKKKKKKKK KKKRG

                (K){20}
 9: LYKLK KKKKKKKKKKKKKKKKKKKK KKRGG

                (K){20}
10: YKLKK KKKKKKKKKKKKKKKKKKKK KRGGP

                (K){20}
11: KLKKK KKKKKKKKKKKKKKKKKKKK RGGPL

               (R,K){20}
12: LKKKK KKKKKKKKKKKKKKKKKKKR GGPLK

AA004873  ck: 7719  len: 66    ! Aao04873 Human polypeptide SEQ ID NO 18765.
1             (R,K){20,20}
                (K){20}
21: SFLIE KKKKKKKKKKKKKKKKKKKK KGGGP

                (K){20}
22: FLIEK KKKKKKKKKKKKKKKKKKKK GGGPL

AA004874  ck: 5753  len: 58    ! Aao04874 Human polypeptide SEQ ID NO 18766.
1             (R,K){20,20}
                (K){20}
10: SEKLP KKKKKKKKKKKKKKKKKKKK KKGKK

                (K){20}
11: EKLPK KKKKKKKKKKKKKKKKKKKK KGKKI

                (K){20}
12: KLPKK KKKKKKKKKKKKKKKKKKKK GKKIK

AA004881  ck: 8841  len: 115   ! Aao04881 Human polypeptide SEQ ID NO 18773.
1             (R,K){20,20}
                (K){20}
24: ITPHP KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
25: TPHPK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
26: PHPKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
27: HPKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
28: PKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGA

                (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KGGAL

                (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK GGALK

AA004917  ck: 3999  len: 103   ! Aao04917 Human polypeptide SEQ ID NO 18809.
1             (R,K){20,20}
                (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK GGGPL

AA004928  ck: 23    len: 43    ! Aao04928 Human polypeptide SEQ ID NO 18820.
1             (R,K){20,20}
                (K){20}
19: FKKKE KKKKKKKKKKKKKKKKKKKK GGGGF
```

```
AAO04969   ck: 4345   len: 57    ! Aao04969 Human polypeptide SEQ ID NO 18861.
1                (R,K){20,20}
                   (K){20}
26: LMGTS KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
27: MGTSK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
28: GTSKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
29: TSKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
30: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                   (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                   (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                   (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK K

                   (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK

AAO05000   ck: 5847   len: 38    ! Aao05000 Human polypeptide SEQ ID NO 18892.
1                (R,K){20,20}
                   (K){20}
 5:  SGRS KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 6: SGRSK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 7: GRSKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 8: RSKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
 9: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                   (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                   (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGG
```

```
                   (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGR

                   (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK RGGR

                 (R,K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKR GGR

AAO05081   ck: 2260   len: 40    ! Aao05081 Human polypeptide SEQ ID NO 18973.
1                (R,K){20,20}
                   (K){20}
 9: SRASP KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
10: RASPK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
11: ASPKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
12: SPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
13: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKS

                   (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKSS

                   (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKSSK

                   (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KSSK

                   (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK SSK

AAO05130   ck: 6597   len: 116   ! Aao05130 Human polypeptide SEQ ID NO 19022.
1                (R,K){20,20}
                   (K){20}
14: ETPFQ KKKKKKKKKKKKKKKKKKKK KRGGG

                   (K){20}
15: TPFQK KKKKKKKKKKKKKKKKKKKK RGGGF

                 (R,K){20}
16: PFQKK KKKKKKKKKKKKKKKKKKKR GGGFL

AAO05191   ck: 4399   len: 85    ! Aao05191 Human polypeptide SEQ ID NO 19083.
1                (R,K){20,20}
                 (R,K){20}
30: GYIQL KKKKKKKKKKKKKKKKKKKR GGAFK

AAO05260   ck: 3618   len: 55    ! Aao05260 Human polypeptide SEQ ID NO 19152.
1                (R,K){20,20}
                   (K){20}
15: PHKHQ KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
16: HKHQK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
 17: KHQKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 18: HQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 19: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
 20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
 21: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

              (K){20}
 22: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGD

              (K){20}
 23: KKKKK KKKKKKKKKKKKKKKKKKKK GGGDF

 AA005369  ck: 1417  len: 70    ! Aao05369 Human polypeptide SEQ ID NO 19261.
1           (R,K){20,20}
              (K){20}
 22: KYHET KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 23: YHETK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 24: HETKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 25: ETKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 26: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

              (K){20}
 27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

              (K){20}
 28: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGE

              (K){20}
 29: KKKKK KKKKKKKKKKKKKKKKKKKK KRGEK

              (K){20}
 30: KKKKK KKKKKKKKKKKKKKKKKKKK RGEKK

             (R,K){20}
 31: KKKKK KKKKKKKKKKKKKKKKKKKR GEKKK

 AA005372  ck: 9880  len: 43    ! Aao05372 Human polypeptide SEQ ID NO 19264.
1           (R,K){20,20}
              (K){20}
  1:       KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKD

              (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKDS

              (K){20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKK KKDSG

              (K){20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKK KDSGG

              (K){20}
 17: KKKKK KKKKKKKKKKKKKKKKKKKK DSGGG

 AA005384  ck: 715  len: 23    ! Aao05384 Human polypeptide SEQ ID NO 19276.
1           (R,K){20,20}
              (K){20}
  4:   LFS KKKKKKKKKKKKKKKKKKKK

 AA005499  ck: 7813  len: 76    ! Aao05499 Human polypeptide SEQ ID NO 19391.
1           (R,K){20,20}
              (K){20}
  7: VQPQQ KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  8: QPQQK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  9: PQQKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 10: QQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 11: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
             (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKA

             (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKAS

             (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKASP

             (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KASPG

             (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK ASPGG

AAO05500   ck: 7443   len: 53    ! Aao05500 Human polypeptide SEQ ID NO 19392.

1          (R,K){20,20}
             (K){20}
 5:  QPQS KKKKKKKKKKKKKKKKKKKK KKASQ

             (K){20}
 6: QPQSK KKKKKKKKKKKKKKKKKKKK KASQK

             (K){20}
 7: PQSKK KKKKKKKKKKKKKKKKKKKK ASQKK

AAO05530   ck: 7354   len: 75    ! Aao05530 Human polypeptide SEQ ID NO 19422.

1          (R,K){20,20}
             (K){20}
11: TTYFL KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
12: TYFLK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
13: YFLKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
14: FLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
15: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKA

             (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKAS

             (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKASS

             (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KASSP

             (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK ASSPP
```

```
          AAO05665   ck: 432    len: 28     ! Aao05665 Human polypeptide SEQ ID NO 19557.
1                    (R,K){20,20}
                       (K){20}
            6: DFLLQ KKKKKKKKKKKKKKKKKKKK RGG

                      (R,K){20}
            7: FLLQK KKKKKKKKKKKKKKKKKKKR GG

          AAO06186   ck: 998    len: 88     ! Aao06186 Human polypeptide SEQ ID NO 20078.
1                    (R,K){20,20}
                       (K){20}
           28: SLLPX KKKKKKKKKKKKKKKKKKKK KKKKG

                       (K){20}
           29: LLPXK KKKKKKKKKKKKKKKKKKKK KKKGG

                       (K){20}
           30: LPXKK KKKKKKKKKKKKKKKKKKKK KKGGG

                       (K){20}
           31: PXKKK KKKKKKKKKKKKKKKKKKKK KGGGG

                       (K){20}
           32: XKKKK KKKKKKKKKKKKKKKKKKKK GGGGF

          AAO06357   ck: 4679   len: 52     ! Aao06357 Human polypeptide SEQ ID NO 20249.
1                    (R,K){20,20}
                       (K){20}
            9: FTCLI KKKKKKKKKKKKKKKKKKKK KKKIK

                       (K){20}
           10: TCLIK KKKKKKKKKKKKKKKKKKKK KKIKX

                       (K){20}
           11: CLIKK KKKKKKKKKKKKKKKKKKKK KIKXK

                       (K){20}
           12: LIKKK KKKKKKKKKKKKKKKKKKKK IKXKK

          AAO06429   ck: 2585   len: 71     ! Aao06429 Human polypeptide SEQ ID NO 20321.
1                    (R,K){20,20}
                       (K){20}
           29: VIIMX KKKKKKKKKKKKKKKKKKKK KKKKK

                       (K){20}
           30: IIMXK KKKKKKKKKKKKKKKKKKKK KKKKG

                       (K){20}
           31: IMXKK KKKKKKKKKKKKKKKKKKKK KKKGR

                       (K){20}
           32: MXKKK KKKKKKKKKKKKKKKKKKKK KKGRG

                       (K){20}
           33: XKKKK KKKKKKKKKKKKKKKKKKKK KGRGG

                       (K){20}
           34: KKKKK KKKKKKKKKKKKKKKKKKKK GRGGA

          AAO06922   ck: 7296   len: 111    ! Aao06922 Human polypeptide SEQ ID NO 20814.
1                    (R,K){20,20}
                       (K){20}
           13: HSSEL KKKKKKKKKKKKKKKKKKKK GGPPK
```

```
          AAO07241   ck: 3134   len: 40     ! Aao07241 Human polypeptide SEQ ID NO 21133.
1                    (R,K){20,20}
                       (K){20}
            4:   YFP KKKKKKKKKKKKKKKKKKKK RKKKX

                      (R,K){20}
            5:  YFPK KKKKKKKKKKKKKKKKKKKR KKKXK

                      (R,K){20}
            6: YFPKK KKKKKKKKKKKKKKKKKKRK KKXKG

                      (R,K){20}
            7: FPKKK KKKKKKKKKKKKKKKKKRKK KXKGG

                      (R,K){20}
            8: PKKKK KKKKKKKKKKKKKKKKRKKK XKGGG

          AAO07354   ck: 6513   len: 133    ! Aao07354 Human polypeptide SEQ ID NO 21246.
1                    (R,K){20,20}
                       (K){20}
           36: KXYHL KKKKKKKKKKKKKKKKKKKK KKKGG

                       (K){20}
           37: XYHLK KKKKKKKKKKKKKKKKKKKK KKGGP

                       (K){20}
           38: YHLKK KKKKKKKKKKKKKKKKKKKK KGGPL

                       (K){20}
           39: HLKKK KKKKKKKKKKKKKKKKKKKK GGPLK

          AAO07410   ck: 693    len: 80     ! Aao07410 Human polypeptide SEQ ID NO 21302.
1                    (R,K){20,20}
                       (K){20}
           56: HKNPI KKKKKKKKKKKKKKKKKKKK RGGGF

                      (R,K){20}
           57: KNPIK KKKKKKKKKKKKKKKKKKKR GGGF

          AAO07412   ck: 4623   len: 166    ! Aao07412 Human polypeptide SEQ ID NO 21304.
1                    (R,K){20,20}
                       (K){20}
           93: VNTQX KKKKKKKKKKKKKKKKKKKK KKKKK

                       (K){20}
           94: NTQXK KKKKKKKKKKKKKKKKKKKK KKKKK

                       (K){20}
           95: TQXKK KKKKKKKKKKKKKKKKKKKK KKKKK

                       (K){20}
           96: QXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                       (K){20}
           97: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                       (K){20}
           98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                       (K){20}
           99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                       (K){20}
          100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
  101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
  104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
  105: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGE

              (K){20}
  106: KKKKK KKKKKKKKKKKKKKKKKKKK KGGEN

              (K){20}
  107: KKKKK KKKKKKKKKKKKKKKKKKKK GGENN

   AA007505  ck: 3528  len: 72    ! Aao07505 Human polypeptide SEQ ID NO 21397.

1           (R,K){20,20}
              (K){20}
   14: HLVEA KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   15: LVEAK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   16: VEAKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   17: EAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   18: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKD

              (K){20}
   20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKDS

              (K){20}
   21: KKKKK KKKKKKKKKKKKKKKKKKKK KKDSR

              (K){20}
   22: KKKKK KKKKKKKKKKKKKKKKKKKK KDSRG

              (K){20}
   23: KKKKK KKKKKKKKKKKKKKKKKKKK DSRGG

   AA007509  ck: 4998  len: 140   ! Aao07509 Human polypeptide SEQ ID NO 21401.

1           (R,K){20,20}
              (K){20}
   17: KVKXE KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   18: VKXEK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   19: KXEKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   20: XEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   21: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
   29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
   30: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

              (K){20}
   31: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGA

              (K){20}
   32: KKKKK KKKKKKKKKKKKKKKKKKKK GGGAS

   AA007594  ck: 7271  len: 93    ! Aao07594 Human polypeptide SEQ ID NO 21486.

1           (R,K){20,20}
              (K){20}
    9: FLLLG KKKKKKKKKKKKKKKKKKKK KKKKT

              (K){20}
   10: LLLGK KKKKKKKKKKKKKKKKKKKK KKKTK

              (K){20}
   11: LLGKK KKKKKKKKKKKKKKKKKKKK KKTKK

              (K){20}
   12: LGKKK KKKKKKKKKKKKKKKKKKKK KTKKN

              (K){20}
   13: GKKKK KKKKKKKKKKKKKKKKKKKK TKKNE

   AA007607  ck: 2434  len: 72    ! Aao07607 Human polypeptide SEQ ID NO 21499.

1           (R,K){20,20}
              (K){20}
   21: KKKKE KKKKKKKKKKKKKKKKKKKK GGGLF

              (K){20}
   48: GLFXE KKKKKKKKKKKKKKKKKKKK GGGLF

   AA007610  ck: 5270  len: 74    ! Aao07610 Human polypeptide SEQ ID NO 21502.

1           (R,K){20,20}
              (K){20}
   19: TELTI KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
   20: ELTIK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                (K){20}
  21: LTIKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  22: TIKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
  23: IKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
  24: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                (K){20}
  25: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGA

                (K){20}
  26: KKKKK KKKKKKKKKKKKKKKKKKKK GGGAL

  AA007755   ck: 6069   len: 90    ! Aao07755 Human polypeptide SEQ ID NO 21647.

1             (R,K){20,20}
                (K){20}
  35: KNQSS KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  36: NQSSK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  37: QSSKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  38: SSKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
  39: SKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
  40: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

                (K){20}
  41: KKKKK KKKKKKKKKKKKKKKKKKKK KGGPL

                (K){20}
  42: KKKKK KKKKKKKKKKKKKKKKKKKK GGPLL

  AA007762   ck: 5115   len: 41    ! Aao07762 Human polypeptide SEQ ID NO 21654.

1             (R,K){20,20}
                (K){20}
  18: IPSLX KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
  19: PSLXK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
  20: SLXKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
  21: LXKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
  22: XKKKK KKKKKKKKKKKKKKKKKKKK

  AA007775   ck: 9827   len: 90    ! Aao07775 Human polypeptide SEQ ID NO 21667.

1             (R,K){20,20}
                (K){20}
  46: PHXPE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  47: HXPEK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
  48: XPEKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
  49: PEKKK KKKKKKKKKKKKKKKKKKKK KKGGA

                (K){20}
  50: EKKKK KKKKKKKKKKKKKKKKKKKK KGGAF

                (K){20}
  51: KKKKK KKKKKKKKKKKKKKKKKKKK GGAFK

  AA007806   ck: 1959   len: 99    ! Aao07806 Human polypeptide SEQ ID NO 21698.

1             (R,K){20,20}
                (K){20}
  41: KLHCI KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  42: LHCIK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  43: HCIKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  44: CIKKK KKKKKKKKKKKKKKKKKKKK KKKKL

                (K){20}
  45: IKKKK KKKKKKKKKKKKKKKKKKKK KKKLW

                (K){20}
  46: KKKKK KKKKKKKKKKKKKKKKKKKK KKLWG

                (K){20}
  47: KKKKK KKKKKKKKKKKKKKKKKKKK KLWGG

                (K){20}
  48: KKKKK KKKKKKKKKKKKKKKKKKKK LWGGG

  AA007872   ck: 8003   len: 135   ! Aao07872 Human polypeptide SEQ ID NO 21764.

1             (R,K){20,20}
                (K){20}
  52: TKKKQ KKKKKKKKKKKKKKKKKKKK GGGLL

  AA007874   ck: 8007   len: 111   ! Aao07874 Human polypeptide SEQ ID NO 21766.

1             (R,K){20,20}
                (K){20}
  90: EEPSI KKKKKKKKKKKKKKKKKKKK KG

                (K){20}
  91: EPSIK KKKKKKKKKKKKKKKKKKKK G

  AA007876   ck: 8130   len: 66    ! Aao07876 Human polypeptide SEQ ID NO 21768.

1             (R,K){20,20}
                (K){20}
  17: IYSFE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  18: YSFEK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
  19: SFEKK KKKKKKKKKKKKKKKKKKKK KKKKQ
```

```
                (K){20}
20: FEKKK KKKKKKKKKKKKKKKKKKKK KKKQK

                (K){20}
21: EKKKK KKKKKKKKKKKKKKKKKKKK KKQKK

                (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KQKKE

                (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK QKKEN

AA007976   ck: 6471   len: 97     ! Aao07976 Human polypeptide SEQ ID NO 21868.
1             (R,K){20,20}
                (K){20}
43: FPIST KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
44: PISTK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
45: ISTKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
46: STKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
47: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKA

                (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKAS

                (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKASS

                (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KASSG

                (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK ASSGG

AA007981   ck: 9551   len: 81     ! Aao07981 Human polypeptide SEQ ID NO 21873.
1             (R,K){20,20}
                (K){20}
54: KPCIY KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
55: PCIYK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
56: CIYKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
57: IYKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
58: YKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK

AA008018   ck: 9768   len: 48     ! Aao08018 Human polypeptide SEQ ID NO 21910.
1             (R,K){20,20}
               (R,K){20}
24: KISQV RKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
25: ISQVR KKKKKKKKKKKKKKKKKKKK KKKK

                (K){20}
26: SQVRK KKKKKKKKKKKKKKKKKKKK KKK

                (K){20}
27: QVRKK KKKKKKKKKKKKKKKKKKKK KK

                (K){20}
28: VRKKK KKKKKKKKKKKKKKKKKKKK K

                (K){20}
29: RKKKK KKKKKKKKKKKKKKKKKKKK

AA008121   ck: 5381   len: 86     ! Aao08121 Human polypeptide SEQ ID NO 22013.
1             (R,K){20,20}
                (K){20}
20: HFGLL KKKKKKKKKKKKKKKKKKKK GGGAF

AA008232   ck: 782    len: 104    ! Aao08232 Human polypeptide SEQ ID NO 22124.
1             (R,K){20,20}
                (K){20}
49: ISQQN KKKKKKKKKKKKKKKKKKKK KDGGG

                (K){20}
50: SQQNK KKKKKKKKKKKKKKKKKKKK DGGGA

AA008450   ck: 6228   len: 31     ! Aao08450 Human polypeptide SEQ ID NO 22342.
1             (R,K){20,20}
                (K){20}
 5:  NNSA KKKKKKKKKKKKKKKKKKKK KKGGA

                (K){20}
 6: NNSAK KKKKKKKKKKKKKKKKKKKK KGGAA

                (K){20}
 7: NSAKK KKKKKKKKKKKKKKKKKKKK GGAAF

AA008458   ck: 1813   len: 59     ! Aao08458 Human polypeptide SEQ ID NO 22350.
1             (R,K){20,20}
                (K){20}
 8: LGGWE KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
```

```
 9: GGWEK KKKKKKKKKKKKKKKKKKKK KKKKN

           (K){20}
10: GWEKK KKKKKKKKKKKKKKKKKKKK KKKNK

           (K){20}
11: WEKKK KKKKKKKKKKKKKKKKKKKK KKNKK

           (K){20}
12: EKKKK KKKKKKKKKKKKKKKKKKKK KNKKK

           (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK NKKKK

AAO08460  ck: 5334  len: 62    ! Aao08460 Human polypeptide SEQ ID NO 22352.

1          (R,K){20,20}
           (K){20}
22: KEFST KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
23: EFSTK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
24: FSTKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
25: STKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
26: TKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKI

           (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKIK

           (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKIKK

           (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KIKKK

           (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK IKKKG

AAO08466  ck: 8728  len: 107   ! Aao08466 Human polypeptide SEQ ID NO 22358.

1          (R,K){20,20}
           (K){20}
57: FWGGV KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
58: WGGVK KKKKKKKKKKKKKKKKKKKK KKKKR

           (K){20}
59: GGVKK KKKKKKKKKKKKKKKKKKKK KKKRR

           (K){20}
60: GVKKK KKKKKKKKKKKKKKKKKKKK KKRRK

           (K){20}
61: VKKKK KKKKKKKKKKKKKKKKKKKK KRRKG

           (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK RRKGR

           (R,K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKR RKGRK

           (R,K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKRR KGRKE

           (R,K){20}
65: KKKKK KKKKKKKKKKKKKKKKKRRK GRKEG

AAO08580  ck: 5164  len: 62    ! Aao08580 Human polypeptide SEQ ID NO 22472.

1          (R,K){20,20}
           (K){20}
30: PILLL KKKKKKKKKKKKKKKKKKKK GGALK

AAO08591  ck: 900   len: 46    ! Aao08591 Human polypeptide SEQ ID NO 22483.

1          (R,K){20,20}
           (K){20}
14: LLCSQ KKKKKKKKKKKKKKKKKKKK GGAFK

AAO08609  ck: 5151  len: 74    ! Aao08609 Human polypeptide SEQ ID NO 22501.

1          (R,K){20,20}
           (K){20}
18: KCVIL KKKKKKKKKKKKKKKKKKKK KGKKL

           (K){20}
19: CVILK KKKKKKKKKKKKKKKKKKKK GKKLK

AAO08612  ck: 557   len: 46    ! Aao08612 Human polypeptide SEQ ID NO 22504.

1          (R,K){20,20}
           (K){20}
20: TFCIM KKKKKKKKKKKKKKKKKKKK KKKGG

           (K){20}
21: FCIMK KKKKKKKKKKKKKKKKKKKK KKGGA

           (K){20}
22: CIMKK KKKKKKKKKKKKKKKKKKKK KGGAL

           (K){20}
23: IMKKK KKKKKKKKKKKKKKKKKKKK GGAL

AAO08616  ck: 4555  len: 41    ! Aao08616 Human polypeptide SEQ ID NO 22508.

1          (R,K){20,20}
           (K){20}
 6: HCALP KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: CALPK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 8: ALPKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 9: LPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
10: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
             (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

             (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

             (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KK

             (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK K

             (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK

AAO08623   ck: 9243  len: 119   ! Aao08623 Human polypeptide SEQ ID NO 22515.
1           (R,K){20,20}
             (R,K){20}
21: DSKQE KKKKKKKKKKKKKKKKKKKR GGAFK

AAO08624   ck: 5590  len: 65    ! Aao08624 Human polypeptide SEQ ID NO 22516.
1           (R,K){20,20}
             (K){20}
11: EGNNW KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
12: GNNWK KKKKKKKKKKKKKKKKKKKK KKKKW

             (K){20}
13: NNWKK KKKKKKKKKKKKKKKKKKKK KKKWG

             (K){20}
14: NWKKK KKKKKKKKKKKKKKKKKKKK KKWGG

             (K){20}
15: WKKKK KKKKKKKKKKKKKKKKKKKK KWGGA

             (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK WGGAL

AAO08631   ck: 6430  len: 135   ! Aao08631 Human polypeptide SEQ ID NO 22523.
1           (R,K){20,20}
             (K){20}
30: TPSRA KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
31: PSRAK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
32: SRAKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
33: RAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
34: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

             (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

             (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGG

             (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGG

             (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK RGGGP

            (R,K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKR GGGPK

AAO08644   ck: 3202  len: 80    ! Aao08644 Human polypeptide SEQ ID NO 22536.
1           (R,K){20,20}
             (K){20}
 2:     E KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 3:    EK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 4:   EKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 5:  EKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 6: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGG

              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK GGGGP

AAO08653  ck: 6448  len: 63    ! Aao08653 Human polypeptide SEQ ID NO 22545.
1
            (R,K){20,20}
              (K){20}
12: PIKLT KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
13: IKLTK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
14: KLTKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
15: LTKKK KKKKKKKKKKKKKKKKKKKK KKKKR

              (K){20}
16: TKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

              (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGA

              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KRGAP

              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK RGAPL

            (R,K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKR GAPLK

AAO08686  ck: 6936  len: 81    ! Aao08686 Human polypeptide SEQ ID NO 22578.
1
            (R,K){20,20}
              (K){20}
23: HPVNH KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
24: PVNHK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
25: VNHKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
26: NHKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
27: HKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

              (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEG

              (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKEGA

              (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KEGAL

              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK EGALG

AAO08701  ck: 588   len: 46    ! Aao08701 Human polypeptide SEQ ID NO 22593.
1
            (R,K){20,20}
              (K){20}
 7: INANS KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 8: NANSK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 9: ANSKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
10: NSKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
11: SKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

              (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGL

              (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK GGGLL

AAO08707  ck: 9359  len: 48    ! Aao08707 Human polypeptide SEQ ID NO 22599.
1
            (R,K){20,20}
              (K){20}
13: TAVTQ KKKKKKKKKKKKKKKKKKKK KKGGG

              (K){20}
14: AVTQK KKKKKKKKKKKKKKKKKKKK KGGGV

              (K){20}
15: VTQKK KKKKKKKKKKKKKKKKKKKK GGGVL
```

```
          AAO08708  ck: 9654  len: 32    ! Aao08708 Human polypeptide SEQ ID NO 22600.
1                   (R,K){20,20}
                      (K){20}
           8: SMTFP KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           9: MTFPK KKKKKKKKKKKKKKKKKKKK KKKK

                      (K){20}
          10: TFPKK KKKKKKKKKKKKKKKKKKKK KKK

                      (K){20}
          11: FPKKK KKKKKKKKKKKKKKKKKKKK KK

                      (K){20}
          12: PKKKK KKKKKKKKKKKKKKKKKKKK K

                      (K){20}
          13: KKKKK KKKKKKKKKKKKKKKKKKKK

          AAO08717  ck: 4688  len: 99    ! Aao08717 Human polypeptide SEQ ID NO 22609.
1                   (R,K){20,20}
                      (K){20}
          80: SFLLI KKKKKKKKKKKKKKKKKKKK

          AAO08772  ck: 1431  len: 105   ! Aao08772 Human polypeptide SEQ ID NO 22664.
1                   (R,K){20,20}
                      (K){20}
          23: IMNYL KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          24: MNYLK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          25: NYLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          26: YLKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                      (K){20}
          27: LKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                      (K){20}
          28: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                      (K){20}
          29: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGP

                      (K){20}
          30: KKKKK KKKKKKKKKKKKKKKKKKKK GGGPF

          AAO08774  ck: 2529  len: 71    ! Aao08774 Human polypeptide SEQ ID NO 22666.
1                   (R,K){20,20}
                      (K){20}
          16: RTIKL KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          17: TIKLK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          18: IKLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          19: KLKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                      (K){20}
          20: LKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                      (K){20}
          21: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

                      (K){20}
          22: KKKKK KKKKKKKKKKKKKKKKKKKK KGGPP

                      (K){20}
          23: KKKKK KKKKKKKKKKKKKKKKKKKK GGPPF

          AAO08817  ck: 529   len: 28    ! Aao08817 Human polypeptide SEQ ID NO 22709.
1                   (R,K){20,20}
                      (K){20}
           3:    AX KKKKKKKKKKKKKKKKKKKK NIIWG

          AAO08820  ck: 1888  len: 84    ! Aao08820 Human polypeptide SEQ ID NO 22712.
1                   (R,K){20,20}
                      (K){20}
          28: CMTFS KKKKKKKKKKKKKKKKKKKK RGGGF

                     (R,K){20}
          29: MTFSK KKKKKKKKKKKKKKKKKKKR GGGFI

          AAO08841  ck: 3529  len: 74    ! Aao08841 Human polypeptide SEQ ID NO 22733.
1                   (R,K){20,20}
                      (K){20}
           3:    PQ KKKKKKKKKKKKKKKKKKKK KWGGG

                      (K){20}
           4:   PQK KKKKKKKKKKKKKKKKKKKK WGGGF

          AAO08857  ck: 2816  len: 49    ! Aao08857 Human polypeptide SEQ ID NO 22749.
1                   (R,K){20,20}
                      (K){20}
          25: PPPTS KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          26: PPTSK KKKKKKKKKKKKKKKKKKKK KKKK

                      (K){20}
          27: PTSKK KKKKKKKKKKKKKKKKKKKK KKK

                      (K){20}
          28: TSKKK KKKKKKKKKKKKKKKKKKKK KK

                      (K){20}
          29: SKKKK KKKKKKKKKKKKKKKKKKKK K

                      (K){20}
          30: KKKKK KKKKKKKKKKKKKKKKKKKK

          AAO08913  ck: 3132  len: 52    ! Aao08913 Human polypeptide SEQ ID NO 22805.
1                   (R,K){20,20}
                      (K){20}
           9: IIKSF KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          10: IKSFK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          11: KSFKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                 (K){20}
       12: SFKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       13: FKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
       27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKI

                 (K){20}
       28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKII

                 (K){20}
       29: KKKKK KKKKKKKKKKKKKKKKKKKK KKII

                 (K){20}
       30: KKKKK KKKKKKKKKKKKKKKKKKKK KII

                 (K){20}
       31: KKKKK KKKKKKKKKKKKKKKKKKKK II

       AAO08914  ck: 4641  len: 25    ! Aao08914 Human polypeptide SEQ ID NO 22806.
1               (R,K){20,20}
                 (R,K){20}
        6: ILMPX KKKKKKKKKKKKKKKKKKKR

       AAO08943  ck: 5770  len: 75    ! Aao08943 Human polypeptide SEQ ID NO 22835.
1               (R,K){20,20}
                 (K){20}
        8: SEWAA KKKKKKKKKKKKKKKKKKKK KKKIF

                 (K){20}
        9: EWAAK KKKKKKKKKKKKKKKKKKKK KKIFF

                 (K){20}
       10: WAAKK KKKKKKKKKKKKKKKKKKKK KIFFF

                 (K){20}
       11: AAKKK KKKKKKKKKKKKKKKKKKKK IFFFG

       AAO08994  ck: 2695  len: 29    ! Aao08994 Human polypeptide SEQ ID NO 22886.
1               (R,K){20,20}
                 (K){20}
        9: NLPSH KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
       10: LPSHK KKKKKKKKKKKKKKKKKKKK

       AAO08995  ck: 6270  len: 26    ! Aao08995 Human polypeptide SEQ ID NO 22887.
1               (R,K){20,20}
                 (K){20}
        6: MAAPP KKKKKKKKKKKKKKKKKKKK I

       AAO09001  ck: 6400  len: 26    ! Aao09001 Human polypeptide SEQ ID NO 22893.
1               (R,K){20,20}
                 (K){20}
        5:  LTSS KKKKKKKKKKKKKKKKKKKK KK

                 (K){20}
        6: LTSSK KKKKKKKKKKKKKKKKKKKK K

                 (K){20}
        7: TSSKK KKKKKKKKKKKKKKKKKKKK

       AAO09016  ck: 4378  len: 119   ! Aao09016 Human polypeptide SEQ ID NO 22908.
1               (R,K){20,20}
                 (R,K){20}
       10: EEEEE RRKKRRRRRRKRRRRKKKKK ILRQK

       AAO09057  ck: 3712  len: 58    ! Aao09057 Human polypeptide SEQ ID NO 22949.
1               (R,K){20,20}
                 (K){20}
        5:  MILN KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
        6: MILNK KKKKKKKKKKKKKKKKKKKK KKKKG

                 (K){20}
        7: ILNKK KKKKKKKKKKKKKKKKKKKK KKKGG

                 (K){20}
        8: LNKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                 (K){20}
        9: NKKKK KKKKKKKKKKKKKKKKKKKK KGGGP

                 (K){20}
       10: KKKKK KKKKKKKKKKKKKKKKKKKK GGGPF

       AAO09066  ck: 2645  len: 29    ! Aao09066 Human polypeptide SEQ ID NO 22958.
1               (R,K){20,20}
```

```
               (K){20}
    4:   AQQ KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    5:  AQQK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    6: AQQKK KKKKKKKKKKKKKKKKKKKK KKKK

               (K){20}
    7: QQKKK KKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
    8: QKKKK KKKKKKKKKKKKKKKKKKKK KK

               (K){20}
    9: KKKKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
   10: KKKKK KKKKKKKKKKKKKKKKKKKK

   AAO09072   ck: 8432   len: 42     ! Aao09072 Human polypeptide SEQ ID NO 22964.
1            (R,K){20,20}
               (K){20}
   11: LIFTL KKKKKKKKKKKKKKKKKKKK KKKEK

               (K){20}
   12: IFTLK KKKKKKKKKKKKKKKKKKKK KKEKY

               (K){20}
   13: FTLKK KKKKKKKKKKKKKKKKKKKK KEKYE

               (K){20}
   14: TLKKK KKKKKKKKKKKKKKKKKKKK EKYEK

   AAO09077   ck: 9524   len: 87     ! Aao09077 Human polypeptide SEQ ID NO 22969.
1            (R,K){20,20}
               (K){20}
   26: RFLLT KKKKKKKKKKKKKKKKKKKK KKKKW

               (K){20}
   27: FLLTK KKKKKKKKKKKKKKKKKKKK KKKWG

               (K){20}
   28: LLTKK KKKKKKKKKKKKKKKKKKKK KKWGG

               (K){20}
   29: LTKKK KKKKKKKKKKKKKKKKKKKK KWGGG

               (K){20}
   30: TKKKK KKKKKKKKKKKKKKKKKKKK WGGGG

   AAO09162   ck: 3841   len: 100    ! Aao09162 Human polypeptide SEQ ID NO 23054.
1            (R,K){20,20}
               (K){20}
   11: KTLFQ KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   12: TLFQK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   13: LFQKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   14: FQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   15: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
   24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

               (K){20}
   25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

               (K){20}
   26: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGG

               (K){20}
   27: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGG

               (K){20}
   28: KKKKK KKKKKKKKKKKKKKKKKKKK RGGGA

             (R,K){20}
   29: KKKKK KKKKKKKKKKKKKKKKKKKR GGGAF

   AAO09258   ck: 7988   len: 42     ! Aao09258 Human polypeptide SEQ ID NO 23150.
1            (R,K){20,20}
               (K){20}
    7: LLFCP KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
    8: LFCPK KKKKKKKKKKKKKKKKKKKK KKKKG

               (K){20}
    9: FCPKK KKKKKKKKKKKKKKKKKKKK KKKGG

               (K){20}
   10: CPKKK KKKKKKKKKKKKKKKKKKKK KKGGP

               (K){20}
   11: PKKKK KKKKKKKKKKKKKKKKKKKK KGGPI

               (K){20}
   12: KKKKK KKKKKKKKKKKKKKKKKKKK GGPIK

   AAO09269   ck: 7553   len: 106    ! Aao09269 Human polypeptide SEQ ID NO 23161.
1            (R,K){20,20}
             (R,K){20}
   39: NTGML KKKKKKKKKKKKKKKKKKKR GGPFL
```

```
          AAO09457  ck: 9658  len: 32    ! Aao09457 Human polypeptide SEQ ID NO 23349.
1                   (R,K){20,20}
                      (K){20}
           9: ALVPQ KKKKKKKKKKKKKKKKKKKK NIKI

          AAO09819  ck: 4709  len: 80    ! Aao09819 Human polypeptide SEQ ID NO 23711.
1                   (R,K){20,20}
                      (K){20}
          54: CFFVX KKKKKKKKKKKKKKKKKKKK EKGGG

          AAO10447  ck: 6342  len: 50    ! Aao10447 Human polypeptide SEQ ID NO 24339.
1                   (R,K){20,20}
                      (K){20}
          25: RFLKL KKKKKKKKKKKKKKKKKKKK RKKKK

                     (R,K){20}
          26: FLKLK KKKKKKKKKKKKKKKKKKKR KKKKI

                     (R,K){20}
          27: LKLKK KKKKKKKKKKKKKKKKKKRK KKKI

                     (R,K){20}
          28: KLKKK KKKKKKKKKKKKKKKKKRKK KKI

                     (R,K){20}
          29: LKKKK KKKKKKKKKKKKKKKKRKKK KI

                     (R,K){20}
          30: KKKKK KKKKKKKKKKKKKKKRKKKK I

          AAO10451  ck: 3955  len: 82    ! Aao10451 Human polypeptide SEQ ID NO 24343.
1                   (R,K){20,20}
                      (K){20}
          59: SRASP KKKKKKKKKKKKKKKKKKKK ARGG

          AAO10467  ck: 7542  len: 116   ! Aao10467 Human polypeptide SEQ ID NO 24359.
1                   (R,K){20,20}
                      (K){20}
          57: KCEFM KKKKKKKKKKKKKKKKKKKK KIGGG

                      (K){20}
          58: CEFMK KKKKKKKKKKKKKKKKKKKK IGGGA

          AAO10564  ck: 9156  len: 98    ! Aao10564 Human polypeptide SEQ ID NO 24456.
1                   (R,K){20,20}
                      (K){20}
          18: NLLTL KKKKKKKKKKKKKKKKKKKK KKEMP

                      (K){20}
          19: LLTLK KKKKKKKKKKKKKKKKKKKK KEMPV

                      (K){20}
          20: LTLKK KKKKKKKKKKKKKKKKKKKK EMPVK

          AAO10608  ck: 9106  len: 67    ! Aao10608 Human polypeptide SEQ ID NO 24500.
1                   (R,K){20,20}
                      (K){20}
          42: CRLSE KKKKKKKKKKKKKKKKKKKK KKKKD

                      (K){20}
          43: RLSEK KKKKKKKKKKKKKKKKKKKK KKKDS

                      (K){20}
          44: LSEKK KKKKKKKKKKKKKKKKKKKK KKDS

                      (K){20}
          45: SEKKK KKKKKKKKKKKKKKKKKKKK KDS

                      (K){20}
          46: EKKKK KKKKKKKKKKKKKKKKKKKK DS

          AAO10638  ck: 621   len: 56    ! Aao10638 Human polypeptide SEQ ID NO 24530.
1                   (R,K){20,20}
                      (K){20}
          33: CEPQP KKKKKKKKKKKKKKKKKKKK KARG

                      (K){20}
          34: EPQPK KKKKKKKKKKKKKKKKKKKK ARG

          AAO10786  ck: 7349  len: 65    ! Aao10786 Human polypeptide SEQ ID NO 24678.
1                   (R,K){20,20}
                      (K){20}
          24: CPXFS KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          25: PXFSK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          26: XFSKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          27: FSKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          28: SKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKV

                      (K){20}
          32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKVF

                      (K){20}
          33: KKKKK KKKKKKKKKKKKKKKKKKKK KKVFF

                      (K){20}
          34: KKKKK KKKKKKKKKKKKKKKKKKKK KVFFF

                      (K){20}
          35: KKKKK KKKKKKKKKKKKKKKKKKKK VFFFE

          AAO10852  ck: 5147  len: 41    ! Aao10852 Human polypeptide SEQ ID NO 24744.
1                   (R,K){20,20}
                      (K){20}
          20: FSRXM KKKKKKKKKKKKKKKKKKKK KK

                      (K){20}
          21: SRXMK KKKKKKKKKKKKKKKKKKKK K

                      (K){20}
```

```
    22: RXMKK KKKKKKKKKKKKKKKKKKKK

    AAO10853  ck: 444   len: 98    ! Aao10853 Human polypeptide SEQ ID NO 24745.
1              (R,K){20,20}
                 (K){20}
    14: LRAHL KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    15: RAHLK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    16: AHLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    17: HLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    18: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                 (K){20}
    25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                 (K){20}
    26: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGE

                 (K){20}
    27: KKKKK KKKKKKKKKKKKKKKKKKKK KRGEL

                 (K){20}
    28: KKKKK KKKKKKKKKKKKKKKKKKKK RGELX

                (R,K){20}
    29: KKKKK KKKKKKKKKKKKKKKKKKKR GELXK

    AAO10859  ck: 2690  len: 70    ! Aao10859 Human polypeptide SEQ ID NO 24751.
1              (R,K){20,20}
                 (K){20}
    44: XGITE KKKKKKKKKKKKKKKKKKKK KIKPG

                 (K){20}
    45: GITEK KKKKKKKKKKKKKKKKKKKK IKPGG

    AAO10933  ck: 849   len: 69    ! Aao10933 Human polypeptide SEQ ID NO 24825.
1              (R,K){20,20}
                 (K){20}
    33: FINTE KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    34: INTEK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    35: NTEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    36: TEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    37: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKI

                 (K){20}
    42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKIE

                 (K){20}
    43: KKKKK KKKKKKKKKKKKKKKKKKKK KKIER

                 (K){20}
    44: KKKKK KKKKKKKKKKKKKKKKKKKK KIERK

                 (K){20}
    45: KKKKK KKKKKKKKKKKKKKKKKKKK IERKI

    AAO10973  ck: 824   len: 51    ! Aao10973 Human polypeptide SEQ ID NO 24865.
1              (R,K){20,20}
                (R,K){20}
    16: SYFXM RKKKKKKKKKKKKKKKKKKK KQNKK

                 (K){20}
    17: YFXMR KKKKKKKKKKKKKKKKKKKK QNKKD

    AAO10983  ck: 694   len: 51    ! Aao10983 Human polypeptide SEQ ID NO 24875.
1              (R,K){20,20}
                 (K){20}
    24: HPILX KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
    25: PILXK KKKKKKKKKKKKKKKKKKKK KKKKT

                 (K){20}
    26: ILXKK KKKKKKKKKKKKKKKKKKKK KKKTG

                 (K){20}
    27: LXKKK KKKKKKKKKKKKKKKKKKKK KKTGG

                 (K){20}
    28: XKKKK KKKKKKKKKKKKKKKKKKKK KTGG

                 (K){20}
    29: KKKKK KKKKKKKKKKKKKKKKKKKK TGG

    AAO10995  ck: 2382  len: 122   ! Aao10995 Human polypeptide SEQ ID NO 24887.
1              (R,K){20,20}
                 (K){20}
    22: PCKGS KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
               (K){20}
23: CKGSK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
24: KGSKK KKKKKKKKKKKKKKKKKKKK KKKKG

               (K){20}
25: GSKKK KKKKKKKKKKKKKKKKKKKK KKKGG

               (K){20}
26: SKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

               (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGG

               (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK GGGGF

AAO10997   ck: 5315  len: 58    ! Aao10997 Human polypeptide SEQ ID NO 24889.

1             (R,K){20,20}
               (K){20}
29: IXIIQ KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
30: XIIQK KKKKKKKKKKKKKKKKKKKK KKKKR

               (K){20}
31: IIQKK KKKKKKKKKKKKKKKKKKKK KKKRG

               (K){20}
32: IQKKK KKKKKKKKKKKKKKKKKKKK KKRGG

               (K){20}
33: QKKKK KKKKKKKKKKKKKKKKKKKK KRGGP

               (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK RGGPP

              (R,K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKR GGPP

AAO11002   ck: 8638  len: 82    ! Aao11002 Human polypeptide SEQ ID NO 24894.

1             (R,K){20,20}
              (R,K){20}
 5:  FKRS KKKKKKKKKKKKKKKKKKKR GGALL

              (R,K){20}
38: XFKRS KKKKKKKKKKKKKKKKKKKR GGALL

AAO11033   ck: 3780  len: 49    ! Aao11033 Human polypeptide SEQ ID NO 24925.

1             (R,K){20,20}
              (R,K){20}
10: QXKFI RKKKKKKKKKKKKKKKKKKK KLSKK

               (K){20}
11: XKFIR KKKKKKKKKKKKKKKKKKKK LSKKI

AAO11048   ck: 7868  len: 85    ! Aao11048 Human polypeptide SEQ ID NO 24940.

1             (R,K){20,20}
               (K){20}
53: IISDP KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
54: ISDPK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
55: SDPKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
56: DPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
57: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

               (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KK

               (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK

AAO11078   ck: 4186  len: 100   ! Aao11078 Human polypeptide SEQ ID NO 24970.

1             (R,K){20,20}
               (K){20}
39: PPKPD KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
40: PKPDK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
41: KPDKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
42: PDKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
43: DKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

               (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

               (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGG
```

```
              (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGG

              (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK RGGGG

             (R,K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKR GGGGF

AAO11124  ck: 5121  len: 61   ! Aao11124 Human polypeptide SEQ ID NO 25016.
1            (R,K){20,20}
              (K){20}
16: KYSFL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
17: YSFLK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
18: SFLKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
19: FLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
20: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK

AAO11139  ck: 8807  len: 68   ! Aao11139 Human polypeptide SEQ ID NO 25031.
1            (R,K){20,20}
              (K){20}
22: VCRFP KKKKKKKKKKKKKKKKKKKK KMEGG

              (K){20}
23: CRFPK KKKKKKKKKKKKKKKKKKKK MEGGG

AAO11165  ck: 7203  len: 62   ! Aao11165 Human polypeptide SEQ ID NO 25057.
1            (R,K){20,20}
              (K){20}
18: DASMV KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
19: ASMVK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
20: SMVKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
21: MVKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
22: VKKKK KKKKKKKKKKKKKKKKKKKK KKKKA

              (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKAS

              (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKASS

              (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KASSS

              (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK ASSSQ

AAO11209  ck: 6562  len: 96   ! Aao11209 Human polypeptide SEQ ID NO 25101.
1            (R,K){20,20}
              (K){20}
63: LLLAY KKKKKKKKKKKKKKKKKKKK KKPKY

              (K){20}
64: LLAYK KKKKKKKKKKKKKKKKKKKK KPKYL

              (K){20}
```

```
          65: LAYKK KKKKKKKKKKKKKKKKKKKK PKYLS

          AAO11210  ck: 863  len: 70   ! Aao11210 Human polypeptide SEQ ID NO 25102.
1                   (R,K){20,20}
                      (K){20}
          31: DLCLC KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          32: LCLCK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          33: CLCKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          34: LCKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          35: CKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKP

                      (K){20}
          42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKPQ

                      (K){20}
          43: KKKKK KKKKKKKKKKKKKKKKKKKK KKPQG

                      (K){20}
          44: KKKKK KKKKKKKKKKKKKKKKKKKK KPQGG

                      (K){20}
          45: KKKKK KKKKKKKKKKKKKKKKKKKK PQGGG

          AAO11214  ck: 9584  len: 68   ! Aao11214 Human polypeptide SEQ ID NO 25106.
1                   (R,K){20,20}
                      (K){20}
          19: NRVCX KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          20: RVCXK KKKKKKKKKKKKKKKKKKKK KKKKR

                      (K){20}
          21: VCXKK KKKKKKKKKKKKKKKKKKKK KKKRG

                      (K){20}
          22: CXKKK KKKKKKKKKKKKKKKKKKKK KKRGG

                      (K){20}
          23: XKKKK KKKKKKKKKKKKKKKKKKKK KRGGP

                      (K){20}
          24: KKKKK KKKKKKKKKKKKKKKKKKKK RGGPL
```

```
                     (R,K){20}
          25: KKKKK KKKKKKKKKKKKKKKKKKKR GGPLK

          AAO11241  ck: 1100  len: 36   ! Aao11241 Human polypeptide SEQ ID NO 25133.
1                   (R,K){20,20}
                      (K){20}
           8: XQTLT KKKKKKKKKKKKKKKKKKKK KAYTS

                      (K){20}
           9: QTLTK KKKKKKKKKKKKKKKKKKKK AYTSK

          AAO11248  ck: 5013  len: 88   ! Aao11248 Human polypeptide SEQ ID NO 25140.
1                   (R,K){20,20}
                      (K){20}
          51: TCLYX KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          52: CLYXK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          53: LYXKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          54: YXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          55: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                      (K){20}
          56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                      (K){20}
          57: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGA

                      (K){20}
          58: KKKKK KKKKKKKKKKKKKKKKKKKK KGGAF

                      (K){20}
          59: KKKKK KKKKKKKKKKKKKKKKKKKK GGAFF

          AAO11293  ck: 6490  len: 73   ! Aao11293 Human polypeptide SEQ ID NO 25185.
1                   (R,K){20,20}
                      (K){20}
          46: KXDYF KKKKKKKKKKKKKKKKKKKK KKKSG

                      (K){20}
          47: XDYFK KKKKKKKKKKKKKKKKKKKK KKSGG

                      (K){20}
          48: DYFKK KKKKKKKKKKKKKKKKKKKK KSGGG

                      (K){20}
          49: YFKKK KKKKKKKKKKKKKKKKKKKK SGGGA

          AAO11324  ck: 6024  len: 93   ! Aao11324 Human polypeptide SEQ ID NO 25216.
1                   (R,K){20,20}
                     (R,K){20}
          37: ISLSA RKKKKKKKKKKKKKKKKKKK KKKGG

                      (K){20}
          38: SLSAR KKKKKKKKKKKKKKKKKKKK KKGGG

                      (K){20}
          39: LSARK KKKKKKKKKKKKKKKKKKKK KGGGG
```

```
                (K){20}
   40: SARKK KKKKKKKKKKKKKKKKKKKK GGGGV

   AAO11327  ck: 3052  len: 102   ! Aao11327 Human polypeptide SEQ ID NO 25219.
1             (R,K){20,20}
                (K){20}
   20: LSKEL KKKKKKKKKKKKKKKKKKKK KKKLG

                (K){20}
   21: SKELK KKKKKKKKKKKKKKKKKKKK KKLGD

                (K){20}
   22: KELKK KKKKKKKKKKKKKKKKKKKK KLGDE

                (K){20}
   23: ELKKK KKKKKKKKKKKKKKKKKKKK LGDEE

   AAO11342  ck: 7189  len: 106   ! Aao11342 Human polypeptide SEQ ID NO 25234.
1             (R,K){20,20}
                (K){20}
   61: FFKTX KKKKKKKKKKKKKKKKKKKK GGPLK

   AAO11346  ck: 7311  len: 85    ! Aao11346 Human polypeptide SEQ ID NO 25238.
1             (R,K){20,20}
                (K){20}
   62: ANHWE KKKKKKKKKKKKKKKKKKKK SGGG

   AAO11352  ck: 1342  len: 51    ! Aao11352 Human polypeptide SEQ ID NO 25244.
1             (R,K){20,20}
                (K){20}
   25: SIFKX KKKKKKKKKKKKKKKKKKKK KNPPS

                (K){20}
   26: IFKXK KKKKKKKKKKKKKKKKKKKK NPPSL

   AAO11705  ck: 7660  len: 101   ! Aao11705 Human polypeptide SEQ ID NO 25597.
1             (R,K){20,20}
                (K){20}
   17: WAPLQ KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
   18: APLQK KKKKKKKKKKKKKKKKKKKK KKGGA

                (K){20}
   19: PLQKK KKKKKKKKKKKKKKKKKKKK KGGAP

                (K){20}
   20: LQKKK KKKKKKKKKKKKKKKKKKKK GGAPL

   AAO11799  ck: 8286  len: 39    ! Aao11799 Human polypeptide SEQ ID NO 25691.
1             (R,K){20,20}
               (R,K){20}
    7: NPPVS RKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    8: PPVSR KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    9: PVSRK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   10: VSRKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   11: SRKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
   12: RKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                (K){20}
   13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

                (K){20}
   14: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

                (K){20}
   15: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGL

                (K){20}
   16: KKKKK KKKKKKKKKKKKKKKKKKKK GGGL

   AAO11820  ck: 8725  len: 42    ! Aao11820 Human polypeptide SEQ ID NO 25712.
1             (R,K){20,20}
                (K){20}
   17: NFHAL KKKKKKKKKKKKKKKKKKKK SSSLR

   AAO11828  ck: 4807  len: 41    ! Aao11828 Human polypeptide SEQ ID NO 25720.
1             (R,K){20,20}
                (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKR

                (K){20}
    5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                (K){20}
    6: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGG

                (K){20}
    7: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGG

                (K){20}
    8: KKKKK KKKKKKKKKKKKKKKKKKKK RGGGF

               (R,K){20}
    9: KKKKK KKKKKKKKKKKKKKKKKKKR GGGFK

   AAO11831  ck: 1761  len: 29    ! Aao11831 Human polypeptide SEQ ID NO 25723.
1             (R,K){20,20}
                (K){20}
    1:       KKKKKKKKKKKKKKKKKKKK KKKKA

                (K){20}
    2:     K KKKKKKKKKKKKKKKKKKKK KKKAG

                (K){20}
    3:    KK KKKKKKKKKKKKKKKKKKKK KKAGG

                (K){20}
    4:   KKK KKKKKKKKKKKKKKKKKKKK KAGGG
```

```
            (K){20}
 5:   KKKK KKKKKKKKKKKKKKKKKKKK AGGGA

AAO11843  ck: 86    len: 36    ! Aao11843 Human polypeptide SEQ ID NO 25735.

1          (R,K){20,20}
            (K){20}
 1:        KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 2:      K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:     KK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 4:    KKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 5:   KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 6:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 7:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 8:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 9:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
10:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
11:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKS

            (K){20}
12:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKSG

            (K){20}
13:  KKKKK KKKKKKKKKKKKKKKKKKKK KKSG

            (K){20}
14:  KKKKK KKKKKKKKKKKKKKKKKKKK KSG

            (K){20}
15:  KKKKK KKKKKKKKKKKKKKKKKKKK SG

AAO11844  ck: 3574  len: 49    ! Aao11844 Human polypeptide SEQ ID NO 25736.

1          (R,K){20,20}
            (K){20}
18:  ILYME KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
19:  LYMEK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
20:  YMEKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
21:  MEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
22:  EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
23:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKY

            (K){20}
24:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKYS

            (K){20}
25:  KKKKK KKKKKKKKKKKKKKKKKKKK KKYSR

            (K){20}
26:  KKKKK KKKKKKKKKKKKKKKKKKKK KYSR

            (K){20}
27:  KKKKK KKKKKKKKKKKKKKKKKKKK YSR

AAO11845  ck: 8375  len: 45    ! Aao11845 Human polypeptide SEQ ID NO 25737.

1          (R,K){20,20}
            (K){20}
15:  SPSTL KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
16:  PSTLK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
17:  STLKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
18:  TLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
19:  LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
20:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKKT

            (K){20}
21:  KKKKK KKKKKKKKKKKKKKKKKKKK KKKTG

            (K){20}
22:  KKKKK KKKKKKKKKKKKKKKKKKKK KKTG

            (K){20}
23:  KKKKK KKKKKKKKKKKKKKKKKKKK KTG

            (K){20}
24:  KKKKK KKKKKKKKKKKKKKKKKKKK TG

AAO11849  ck: 284   len: 36    ! Aao11849 Human polypeptide SEQ ID NO 25741.

1          (R,K){20,20}
            (K){20}
13:  KCLCE KKKKKKKKKKKKKKKKKKKK KRKR

            (K){20}
14:  CLCEK KKKKKKKKKKKKKKKKKKKK RKR

           (R,K){20}
15:  LCEKK KKKKKKKKKKKKKKKKKKKR KR

           (R,K){20}
16:  CEKKK KKKKKKKKKKKKKKKKKKRK R

           (R,K){20}
17:  EKKKK KKKKKKKKKKKKKKKKKRKR

AAO11886  ck: 5282  len: 97    ! Aao11886 Human polypeptide SEQ ID NO 25778.

1          (R,K){20,20}
            (K){20}
32:  SNSSL KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
33: NSSLK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
34: SSLKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
35: SLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
36: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

              (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGG

              (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK GGGGV

AAO11888  ck: 6488  len: 31   ! Aao11888 Human polypeptide SEQ ID NO 25780.

1           (R,K){20,20}
              (K){20}
 2:       D KKKKKKKKKKKKKKKKKKKK KAGGG

              (K){20}
 3:      DK KKKKKKKKKKKKKKKKKKKK AGGGL

AAO11903  ck: 8876  len: 48   ! Aao11903 Human polypeptide SEQ ID NO 25795.

1           (R,K){20,20}
              (K){20}
17: VIKYV KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
18: IKYVK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
19: KYVKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
20: YVKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
21: VKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK

AAO11999  ck: 3954  len: 47   ! Aao11999 Human polypeptide SEQ ID NO 25891.

1           (R,K){20,20}
              (K){20}
17: IDEGL KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
18: DEGLK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
19: EGLKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
20: GLKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
21: LKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGA

              (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KGGAP

              (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK GGAP

AAO12098  ck: 8434  len: 39   ! Aao12098 Human polypeptide SEQ ID NO 25990.

1           (R,K){20,20}
              (K){20}
13: TCKFG KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
14: CKFGK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
15: KFGKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
16: FGKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
17: GKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KK
```

```
               (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK

AAO12105  ck: 6255  len: 80    ! Aao12105 Human polypeptide SEQ ID NO 25997.
1             (R,K){20,20}
               (K){20}
53: EATSL KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
54: ATSLK KKKKKKKKKKKKKKKKKKKK KKKKA

               (K){20}
55: TSLKK KKKKKKKKKKKKKKKKKKKK KKKAR

               (K){20}
56: SLKKK KKKKKKKKKKKKKKKKKKKK KKARG

               (K){20}
57: LKKKK KKKKKKKKKKKKKKKKKKKK KARG

               (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK ARG

AAO12179  ck: 1295  len: 69    ! Aao12179 Human polypeptide SEQ ID NO 26071.
1             (R,K){20,20}
               (K){20}
35: PKFSV KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
36: KFSVK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
37: FSVKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
38: SVKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
39: VKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

               (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

               (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

               (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGA

               (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK GGGA

AAO12180  ck: 9840  len: 67    ! Aao12180 Human polypeptide SEQ ID NO 26072.
1             (R,K){20,20}
               (K){20}
38: CSYLP KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
39: SYLPK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
40: YLPKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
41: LPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
42: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

               (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KK

               (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK

AAO12187  ck: 4700  len: 60    ! Aao12187 Human polypeptide SEQ ID NO 26079.
1             (R,K){20,20}
               (K){20}
30: SCKKL KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
31: CKKLK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
32: KKLKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
33: KLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
34: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

               (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

               (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KK

               (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK K

               (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK
```

```
        AAO12203   ck: 4083   len: 41    ! Aao12203 Human polypeptide SEQ ID NO 26095.
1                  (R,K){20,20}
                      (K){20}
         6: SWCCL KKKKKKKKKKKKKKKKKKKK KKKPG

                      (K){20}
         7: WCCLK KKKKKKKKKKKKKKKKKKKK KKPGG

                      (K){20}
         8: CCLKK KKKKKKKKKKKKKKKKKKKK KPGGG

                      (K){20}
         9: CLKKK KKKKKKKKKKKKKKKKKKKK PGGGA

        AAO12215   ck: 6903   len: 73    ! Aao12215 Human polypeptide SEQ ID NO 26107.
1                  (R,K){20,20}
                      (K){20}
        50: PPFLP KKKKKKKKKKKKKKKKKKKK KTGG

                      (K){20}
        51: PFLPK KKKKKKKKKKKKKKKKKKKK TGG

        AAO12243   ck: 8474   len: 39    ! Aao12243 Human polypeptide SEQ ID NO 26135.
1                  (R,K){20,20}
                      (K){20}
        11: MISFI KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        12: ISFIK KKKKKKKKKKKKKKKKKKKK KKKKG

                      (K){20}
        13: SFIKK KKKKKKKKKKKKKKKKKKKK KKKGG

                      (K){20}
        14: FIKKK KKKKKKKKKKKKKKKKKKKK KKGGP

                      (K){20}
        15: IKKKK KKKKKKKKKKKKKKKKKKKK KGGPL

                      (K){20}
        16: KKKKK KKKKKKKKKKKKKKKKKKKK GGPL

        AAO12250   ck: 1538   len: 69    ! Aao12250 Human polypeptide SEQ ID NO 26142.
1                  (R,K){20,20}
                      (K){20}
        31: SPSNL KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        32: PSNLK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        33: SNLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        34: NLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        35: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        40: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKR

                      (K){20}
        41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKRG

                      (K){20}
        42: KKKKK KKKKKKKKKKKKKKKKKKKK KKRGG

                      (K){20}
        43: KKKKK KKKKKKKKKKKKKKKKKKKK KRGGG

                      (K){20}
        44: KKKKK KKKKKKKKKKKKKKKKKKKK RGGGA

                    (R,K){20}
        45: KKKKK KKKKKKKKKKKKKKKKKKKR GGGAF

        AAO12274   ck: 2739   len: 37    ! Aao12274 Human polypeptide SEQ ID NO 26166.
1                  (R,K){20,20}
                      (K){20}
        14: KGLIN KKKKKKKKKKKKKKKKKKKK KKKG

                      (K){20}
        15: GLINK KKKKKKKKKKKKKKKKKKKK KKG

                      (K){20}
        16: LINKK KKKKKKKKKKKKKKKKKKKK KG

                      (K){20}
        17: INKKK KKKKKKKKKKKKKKKKKKKK G

        AAO12280   ck: 7481   len: 66    ! Aao12280 Human polypeptide SEQ ID NO 26172.
1                  (R,K){20,20}
                      (K){20}
        29: HIPPL KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        30: IPPLK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        31: PPLKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        32: PLKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        33: LKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        36: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        37: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
        38: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKS
```

```
             (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKK KKKSG

             (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKK KKSGG

             (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKK KSGGG

             (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKK SGGGA

AAO12447   ck: 5017   len: 47    ! Aao12447 Human polypeptide SEQ ID NO 26339.
1          (R,K){20,20}
             (K){20}
14: HLVCE KKKKKKKKKKKKKKKKKKKK KRGGA

             (K){20}
15: LVCEK KKKKKKKKKKKKKKKKKKKK RGGAL

            (R,K){20}
16: VCEKK KKKKKKKKKKKKKKKKKKKR GGALK

AAO12476   ck: 837   len: 104    ! Aao12476 Human polypeptide SEQ ID NO 26368.
1          (R,K){20,20}
             (K){20}
65: RFCHQ KKKKKKKKKKKKKKKKKKKK KGFFF

             (K){20}
66: FCHQK KKKKKKKKKKKKKKKKKKKK GFFFW

AAO12548   ck: 6973   len: 74    ! Aao12548 Human polypeptide SEQ ID NO 26440.
1          (R,K){20,20}
             (K){20}
38: AVLPL KKKKKKKKKKKKKKKKKKKK KKGGG

             (K){20}
39: VLPLK KKKKKKKKKKKKKKKKKKKK KGGGV

             (K){20}
40: LPLKK KKKKKKKKKKKKKKKKKKKK GGGVF

AAO12553   ck: 1903   len: 33    ! Aao12553 Human polypeptide SEQ ID NO 26445.
1          (R,K){20,20}
             (K){20}
 9: SCCFI KKKKKKKKKKKKKKKKKKKK KRGAP

             (K){20}
10: CCFIK KKKKKKKKKKKKKKKKKKKK RGAP

            (R,K){20}
11: CFIKK KKKKKKKKKKKKKKKKKKKR GAP

AAO13164   ck: 2798   len: 71    ! Aao13164 Human polypeptide SEQ ID NO 27056.
1          (R,K){20,20}
            (R,K){20}
35: RPPLX RKKKKKKKKKKKKKKKKKKK KEMFK

             (K){20}
36: PPLXR KKKKKKKKKKKKKKKKKKKK EMFKR

AAO13576   ck: 4846   len: 99    ! Aao13576 Human polypeptide SEQ ID NO 27468.
1          (R,K){20,20}
             (K){20}
53: TNNLI KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
54: NNLIK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
55: NLIKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
56: LIKKK KKKKKKKKKKKKKKKKKKKK KKKKG

             (K){20}
57: IKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

             (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGG

             (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KGGGP

             (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK GGGPF

AAO13785   ck: 6241   len: 100    ! Aao13785 Human polypeptide SEQ ID NO 27677.
1          (R,K){20,20}
             (K){20}
41: LYAPP KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
42: YAPPK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
43: APPKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
44: PPKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
45: PKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKA

             (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKAG

             (K){20}
```

```
     55: KKKKK KKKKKKKKKKKKKKKKKKKK KKAGG

                (K){20}
     56: KKKKK KKKKKKKKKKKKKKKKKKKK KAGGG

                (K){20}
     57: KKKKK KKKKKKKKKKKKKKKKKKKK AGGGG

     AAU17983   ck: 7016   len: 315    ! Aau17983 Human immunoglobulin polypeptide S
1             (R,K){20,20}
               (R,K){20}
    273: QVFAP RKKKKKKKKKKKKKKKKKKK KGGRS

                (K){20}
    274: VFAPR KKKKKKKKKKKKKKKKKKKK GGRSR

     AAU18049   ck: 7611   len: 315    ! Aau18049 Human immunoglobulin polypeptide S
1             (R,K){20,20}
               (R,K){20}
    273: QVFAP RKKKKKKKKKKKKKKKKKKK KGGRS

                (K){20}
    274: VFAPR KKKKKKKKKKKKKKKKKKKK GGRSR

     AAM14119   ck: 5383   len: 86     ! Aam14119 Peptide #553 encoded by probe for
1             (R,K){20,20}
               (R,K){20}
     15: RRRRG RRRRRKKKKKKRKKKKKRKK KKRRR

               (R,K){20}
     16: RRRGR RRRRKKKKKKRKKKKKRKKK KRRRR

               (R,K){20}
     17: RRGRR RRRKKKKKKRKKKKKRKKKK RRRRR

               (R,K){20}
     18: RGRRR RRKKKKKKRKKKKKRKKKKR RRRRR

               (R,K){20}
     19: GRRRR RKKKKKKRKKKKKRKKKKRR RRRRR

               (R,K){20}
     20: RRRRR KKKKKKRKKKKKRKKKKRRR RRRRR

               (R,K){20}
     21: RRRRK KKKKKRKKKKKRKKKKRRRR RRRRR

               (R,K){20}
     22: RRRKK KKKKRKKKKKRKKKKRRRRR RRRRR

               (R,K){20}
     23: RRKKK KKKRKKKKKRKKKKRRRRRR RRRRR

               (R,K){20}
     24: RKKKK KKRKKKKKRKKKKRRRRRRR RRRRR

               (R,K){20}
     25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

               (R,K){20}
     26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

               (R,K){20}
     27: KKKKR KKKKKRKKKKRRRRRRRRRR RRRRR

               (R,K){20}
     28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

               (R,K){20}
     29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

               (R,K){20}
     30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

               (R,K){20}
     31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

               (R,K){20}
     32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

               (R,K){20}
     33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

               (R,K){20}
     34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

               (R,K){20}
     35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

               (R,K){20}
     36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
     37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
     38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
     39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
     40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                (R){20}
     41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                (R){20}
     42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

                (R){20}
     43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

                (R){20}
     44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

                (R){20}
     45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

     AAM14961   ck: 1334   len: 86     ! Aam14961 Peptide #1395 encoded by probe for
1             (R,K){20,20}
               (R,K){20}
     57: EEEEG RRRKKKKKRKRKKKKKKKKK KKKKK

               (R,K){20}
     58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

               (R,K){20}
     59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

               (R,K){20}
     60: EGRRR KKKKRKRKKKKKKKKKKKKK KKKKK

               (R,K){20}
     61: GRRRK KKKRKRKKKKKKKKKKKKKK KKKKK

               (R,K){20}
     62: RRRKK KKRKRKKKKKKKKKKKKKKK KKKKK
```

```
          (R,K){20}
63: RRKKK KRKRKKKKKKKKKKKKKKKK KKKK

          (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKKK KKK

          (R,K){20}
65: KKKKR KRKKKKKKKKKKKKKKKKKK KK

          (R,K){20}
66: KKKRK RKKKKKKKKKKKKKKKKKKK K

           (K){20}
67: KKRKR KKKKKKKKKKKKKKKKKKKK


AAM15038  ck: 9082  len: 167   ! Aam15038 Peptide #1472 encoded by probe for

1        (R,K){20,20}
          (R,K){20}
33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

          (R,K){20}
34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

          (R,K){20}
35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR


AAM15527  ck: 2276  len: 89    ! Aam15527 Peptide #1961 encoded by probe for

1        (R,K){20,20}
          (R,K){20}
23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

           (K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

           (K){20}
47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

          (R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

          (R,K){20}
49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE


AAM15826  ck: 1939  len: 130   ! Aam15826 Peptide #2260 encoded by probe for

1        (R,K){20,20}
          (R,K){20}
42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR

          (R,K){20}
43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

          (R,K){20}
44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK

          (R,K){20}
45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

          (R,K){20}
46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

          (R,K){20}
47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

          (R,K){20}
48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

          (R,K){20}
49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

          (R,K){20}
50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

          (R,K){20}
51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

          (R,K){20}
52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

          (R,K){20}
53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE

          (R,K){20}
54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE


AAM16123  ck: 3607  len: 88    ! Aam16123 Peptide #2557 encoded by probe for

1        (R,K){20,20}
           (K){20}
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
                  (K){20}
        56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
        65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

                  (K){20}
        66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

                  (K){20}
        67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

                  (K){20}
        68: KKKKK KKKKKKKKKKKKKKKKKKKK K

                  (K){20}
        69: KKKKK KKKKKKKKKKKKKKKKKKKK

        AAM16691  ck: 3937  len: 85    ! Aam16691 Peptide #3125 encoded by probe for

1                (R,K){20,20}
                  (K){20}
         1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         3:    KK KKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
         7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

                  (K){20}
         8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

                  (K){20}
         9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

                  (K){20}
        10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

                  (K){20}
        11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

        AAM17388  ck: 2686  len: 71    ! Aam17388 Peptide #3822 encoded by probe for

1                (R,K){20,20}
                  (K){20}
        20: KKKKQ KKKKKKKKKKKKKKKKKKKK KKKKR

                  (K){20}
        21: KKKQK KKKKKKKKKKKKKKKKKKKK KKKRK

                  (K){20}
        22: KKQKK KKKKKKKKKKKKKKKKKKKK KKRKR

                  (K){20}
        23: KQKKK KKKKKKKKKKKKKKKKKKKK KRKRK

                  (K){20}
        24: QKKKK KKKKKKKKKKKKKKKKKKKK RKRKK

                  (R,K){20}
        25: KKKKK KKKKKKKKKKKKKKKKKKKR KRKKK

                  (R,K){20}
        26: KKKKK KKKKKKKKKKKKKKKKKKRK RKKKK

                  (R,K){20}
        27: KKKKK KKKKKKKKKKKKKKKKKRKR KKKKK

                  (R,K){20}
        28: KKKKK KKKKKKKKKKKKKKKKRKRK KKKKK

                  (R,K){20}
        29: KKKKK KKKKKKKKKKKKKKKRKRKK KKKKK

                  (R,K){20}
        30: KKKKK KKKKKKKKKKKKKKRKRKKK KKKKK

                  (R,K){20}
        31: KKKKK KKKKKKKKKKKKKRKRKKKK KKKKK

                  (R,K){20}
        32: KKKKK KKKKKKKKKKKKRKRKKKKK KKKKK

                  (R,K){20}
        33: KKKKK KKKKKKKKKKKRKRKKKKKK KKKKK

                  (R,K){20}
        34: KKKKK KKKKKKKKKKRKRKKKKKKK KKKKK

                  (R,K){20}
        35: KKKKK KKKKKKKKKRKRKKKKKKKK KKKKK

                  (R,K){20}
        36: KKKKK KKKKKKKKRKRKKKKKKKKK KKKKK

                  (R,K){20}
        37: KKKKK KKKKKKKRKRKKKKKKKKKK KKKKK

                  (R,K){20}
        38: KKKKK KKKKKKRKRKKKKKKKKKKK KKKKK

                  (R,K){20}
        39: KKKKK KKKKKRKRKKKKKKKKKKKK KKKKK

                  (R,K){20}
        40: KKKKK KKKKRKRKKKKKKKKKKKKK KKKKK
```

```
                (R,K){20}
    41: KKKKK KKKRKRKKKKKKKKKKKKKK KKKKK

                (R,K){20}
    42: KKKKK KKRKRKKKKKKKKKKKKKKK KKKKK

                (R,K){20}
    43: KKKKK KRKRKKKKKKKKKKKKKKKK KKKKK

                (R,K){20}
    44: KKKKK RKRKKKKKKKKKKKKKKKKK KKKKK

                (R,K){20}
    45: KKKKR KRKKKKKKKKKKKKKKKKKK KKKKS

                (R,K){20}
    46: KKKRK RKKKKKKKKKKKKKKKKKKK KKKSA

                 (K){20}
    47: KKRKR KKKKKKKKKKKKKKKKKKKK KKSAH

                 (K){20}
    48: KRKRK KKKKKKKKKKKKKKKKKKKK KSAH

                 (K){20}
    49: RKRKK KKKKKKKKKKKKKKKKKKKK SAH

    AAM18532   ck: 1560   len: 88    ! Aam18532 Peptide #4966 encoded by probe for

1              (R,K){20,20}
                 (R){20}
    43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                 (R){20}
    58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

                 (R){20}
    59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

                 (R){20}
    60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

                 (R){20}
    61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

                 (R){20}
    62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

    AAM20089   ck: 8343   len: 66    ! Aam20089 Peptide #6523 encoded by probe for

1              (R,K){20,20}
                (R,K){20}
     6: ETERE KKKKRKKKKKKKKKKKKKKK KKKNK

                (R,K){20}
     7: TEREK KKKRKKKKKKKKKKKKKKKK KKNKK

                (R,K){20}
     8: EREKK KKRKKKKKKKKKKKKKKKKK KNKKK

                (R,K){20}
     9: REKKK KRKKKKKKKKKKKKKKKKKK NKKKK

    AAM21941   ck: 4895   len: 51    ! Aam21941 Peptide #8375 encoded by probe for

1              (R,K){20,20}
                (R,K){20}
    18: LFKPM RKRRRKKKRRRKKRKKKRRR KKLTT

                (R,K){20}
    19: FKPMR KRRRKKKRRRKKRKKKRRRK KLTTT

                (R,K){20}
    20: KPMRK RRRKKKRRRKKRKKKRRRKK LTTTT

    AAM26526   ck: 5383   len: 86    ! Aam26526 Peptide #563 encoded by probe for

1              (R,K){20,20}
                (R,K){20}
    15: RRRRG RRRRRKKKKKRKKKKKRKK KKRRR

                (R,K){20}
    16: RRRGR RRRRKKKKKRKKKKKRKKK KRRRR

                (R,K){20}
    17: RRGRR RRRKKKKKRKKKKKRKKKK RRRRR

                (R,K){20}
    18: RGRRR RRKKKKKRKKKKKRKKKKR RRRRR

                (R,K){20}
    19: GRRRR RKKKKKRKKKKKRKKKKRR RRRRR

                (R,K){20}
    20: RRRRR KKKKKRKKKKKRKKKKRRR RRRRR

                (R,K){20}
    21: RRRRK KKKKRKKKKKRKKKKRRRR RRRRR

                (R,K){20}
```

```
22: RRRKK KKKKRKKKKKRKKKKRRRRRR RRRRR

          (R,K){20}
23: RRKKK KKKRKKKKKRKKKKRRRRRR RRRRR

          (R,K){20}
24: RKKKK KKRKKKKKRKKKKRRRRRRR RRRRR

          (R,K){20}
25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

          (R,K){20}
26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

          (R,K){20}
27: KKKKR KKKKKRKKKKRRRRRRRRRR RRRRR

          (R,K){20}
28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

          (R,K){20}
29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

          (R,K){20}
30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

          (R,K){20}
31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

          (R,K){20}
32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

          (R,K){20}
33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

          (R,K){20}
34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

          (R,K){20}
35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

          (R,K){20}
36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

           (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

           (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

           (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

           (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

AAM27395  ck: 1334  len: 86   ! Aam27395 Peptide #1432 encoded by probe for

1           (R,K){20,20}
             (R,K){20}
57: EEEEG RRRKKKKRKRKKKKKKKKKK KKKKK

             (R,K){20}
58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

             (R,K){20}
59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

             (R,K){20}
60: EGRRR KKKKRKRKKKKKKKKKKKKK KKKKK

             (R,K){20}
61: GRRRK KKKRKRKKKKKKKKKKKKKK KKKKK

             (R,K){20}
62: RRRKK KKRKRKKKKKKKKKKKKKKK KKKKK

             (R,K){20}
63: RRKKK KRKRKKKKKKKKKKKKKKKK KKKK

             (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKKK KKK

             (R,K){20}
65: KKKKR KRKKKKKKKKKKKKKKKKKK KK

             (R,K){20}
66: KKKRK RKKKKKKKKKKKKKKKKKKK K

              (K){20}
67: KKRKR KKKKKKKKKKKKKKKKKKKK

AAM27479  ck: 9082  len: 167   ! Aam27479 Peptide #1516 encoded by probe for

1           (R,K){20,20}
             (R,K){20}
33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

             (R,K){20}
34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

             (R,K){20}
35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR

AAM28016  ck: 2276  len: 89   ! Aam28016 Peptide #2053 encoded by probe for

1           (R,K){20,20}
             (R,K){20}
23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

              (K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

              (K){20}
47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

             (R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

             (R,K){20}
49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

AAM28336  ck: 1939  len: 130   ! Aam28336 Peptide #2373 encoded by probe for

1           (R,K){20,20}
             (R,K){20}
42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR
```

```
          (R,K){20}
43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

          (R,K){20}
44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK

          (R,K){20}
45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

          (R,K){20}
46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

          (R,K){20}
47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

          (R,K){20}
48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

          (R,K){20}
49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

          (R,K){20}
50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

          (R,K){20}
51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

          (R,K){20}
52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

          (R,K){20}
53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE

          (R,K){20}
54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE

AAM28616  ck: 3607  len: 88   ! Aam28616 Peptide #2653 encoded by probe for

1         (R,K){20,20}
            (K){20}
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

            (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

            (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

            (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK K

            (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK

AAM29178  ck: 3937  len: 85   ! Aam29178 Peptide #3215 encoded by probe for

1         (R,K){20,20}
            (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

          (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

          (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

          (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

          (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

AAM30989  ck: 1560  len: 88    ! Aam30989 Peptide #5026 encoded by probe for

1        (R,K){20,20}
          (R){20}
43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

          (R){20}
59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

          (R){20}
60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

          (R){20}
61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

          (R){20}
62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

AAM31173  ck: 2324  len: 36    ! Aam31173 Peptide #5210 encoded by probe for

1        (R,K){20,20}
          (R,K){20}
16: KERKT KKRKRKRRRRRKKRRRKKRR R

          (R,K){20}
17: ERKTK KRKRKRRRRRKKRRRKKRRR

AAM33953  ck: 8343  len: 66    ! Aam33953 Peptide #7990 encoded by probe for

1        (R,K){20,20}
          (R,K){20}
 6: ETERE KKKKRKKKKKKKKKKKKKKK KKKNK

          (R,K){20}
 7: TEREK KKKRKKKKKKKKKKKKKKKK KKNKK

          (R,K){20}
 8: EREKK KKRKKKKKKKKKKKKKKKKK KNKKK

          (R,K){20}
 9: REKKK KRKKKKKKKKKKKKKKKKKK NKKKK

AAM36456  ck: 2394  len: 57    ! Aam36456 Peptide #10493 encoded by probe fo

1        (R,K){20,20}
          (R){20}
20: EEEEG RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
21: EEEGR RRRRRRRRRRRRRRRRRRRR RRRRG

          (R){20}
22: EEGRR RRRRRRRRRRRRRRRRRRRR RRRGR

          (R){20}
23: EGRRR RRRRRRRRRRRRRRRRRRRR RRGRR

          (R){20}
24: GRRRR RRRRRRRRRRRRRRRRRRRR RGRRR

          (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRRRR GRRRR

AAM37018  ck: 4228  len: 24    ! Aam37018 Peptide #11055 encoded by probe fo

1        (R,K){20,20}
          (R,K){20}
 1:       RRRRRRRKKKKRRRRRRRRR RKRT

          (R,K){20}
 2:     R RRRRRRKKKKRRRRRRRRRR KRT
```

```
              (R,K){20}
 3:     RR RRRRRKKKKRRRRRRRRRRK RT

              (R,K){20}
 4:    RRR RRRRKKKKRRRRRRRRRRKR T

AAM37794  ck: 3301  len: 52    ! Aam37794 Peptide #11831 encoded by probe fo

1            (R,K){20,20}
              (R,K){20}
12: KKKKN KKKKKKKKKRKRKRKRKKKKK KKKKK

              (R,K){20}
13: KKKNK KKKKKKKKRKRKRKRKKKKKK KKKKK

              (R,K){20}
14: KKNKK KKKKKKKRKRKRKRKKKKKKK KKKKK

              (R,K){20}
15: KNKKK KKKKKKRKRKRKRKKKKKKKK KKKKK

              (R,K){20}
16: NKKKK KKKKKRKRKRKRKKKKKKKKK KKKKK

              (R,K){20}
17: KKKKK KKKKRKRKRKRKKKKKKKKKK KKKKK

              (R,K){20}
18: KKKKK KKKRKRKRKRKKKKKKKKKKK KKKKK

              (R,K){20}
19: KKKKK KKRKRKRKRKKKKKKKKKKKK KKKKK

              (R,K){20}
20: KKKKK KRKRKRKRKKKKKKKKKKKKK KKKKK

              (R,K){20}
21: KKKKK RKRKRKRKKKKKKKKKKKKKK KKKKK

              (R,K){20}
22: KKKKR KRKRKRKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
23: KKKRK RKRKRKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
24: KKRKR KRKRKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
25: KRKRK RKRKKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
26: RKRKR KRKKKKKKKKKKKKKKKKKKK KKKKK

              (R,K){20}
27: KRKRK RKKKKKKKKKKKKKKKKKKKK KKKKA

               (K){20}
28: RKRKR KKKKKKKKKKKKKKKKKKKK KKKAF

               (K){20}
29: KRKRK KKKKKKKKKKKKKKKKKKKK KKAF

               (K){20}
30: RKRKK KKKKKKKKKKKKKKKKKKKK KAF

               (K){20}
31: KRKKK KKKKKKKKKKKKKKKKKKKK AF

AAM38273  ck: 4895  len: 51    ! Aam38273 Peptide #12310 encoded by probe fo

1            (R,K){20,20}
              (R,K){20}
18: LFKPM RKRRRKKKRRRKKRKKKRRR KKLTT

              (R,K){20}
19: FKPMR KRRRKKKRRRKKRKKKRRRK KLTTT

              (R,K){20}
20: KPMRK RRRKKKRRRKKRKKKRRRKK LTTTT

AAU04283  ck: 8137  len: 45    ! Aau04283 Trimeric fusogenic peptide #2 used

1            (R,K){20,20}
               (K){20}
 4:    YKA KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 5:   YKAK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 6: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 7: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 8: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

               (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

               (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

               (K){20}
```

```
          23: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

                      (K){20}
          24: KKKKK KKKKKKKKKKKKKKKKKKKK WK

          AAU04285   ck: 4361   len: 59    ! Aau04285 Nuclear ligand #2 used in nucleic

1                   (R,K){20,20}
                      (K){20}
          18: APYKA KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          19: PYKAK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          20: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          21: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          22: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

                      (K){20}
          35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

                      (K){20}
          36: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

                      (K){20}
          37: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

                      (K){20}
          38: KKKKK KKKKKKKKKKKKKKKKKKKK WK
```

```
          AAU04287   ck: 4925   len: 100   ! Aau04287 Poly-L-Lysine used in nucleic acid

1                   (R,K){20,20}
                      (K){20}
           1:       KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
           9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
          23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                      (K){20}
```

```
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
```

```
         (K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
74: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

         (K){20}
77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

         (K){20}
78: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

         (K){20}
79: KKKKK KKKKKKKKKKKKKKKKKKKK KK

         (K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKKK K

         (K){20}
81: KKKKK KKKKKKKKKKKKKKKKKKKK

AAM01857  ck: 5383  len: 86    ! Aam01857 Peptide #539 encoded by probe for

1          (R,K){20,20}
            (R,K){20}
15: RRRRG RRRRRKKKKKRKKKKKRKK KKRRR

            (R,K){20}
16: RRRGR RRRRKKKKKKRKKKKKRKKK KRRRR

            (R,K){20}
17: RRGRR RRRKKKKKKRKKKKKRKKKK RRRRR

            (R,K){20}
18: RGRRR RRKKKKKKRKKKKKRKKKKR RRRRR

            (R,K){20}
19: GRRRR RKKKKKKRKKKKKRKKKKRR RRRRR

            (R,K){20}
20: RRRRR KKKKKKRKKKKKRKKKKRRR RRRRR

            (R,K){20}
21: RRRRK KKKKKRKKKKKRKKKKRRRR RRRRR

            (R,K){20}
22: RRRKK KKKKRKKKKKRKKKKRRRRR RRRRR

            (R,K){20}
23: RRKKK KKKRKKKKKRKKKKRRRRRR RRRRR

            (R,K){20}
24: RKKKK KKRKKKKKRKKKKRRRRRRR RRRRR

            (R,K){20}
25: KKKKK KRKKKKKRKKKKRRRRRRRR RRRRR

            (R,K){20}
26: KKKKK RKKKKKRKKKKRRRRRRRRR RRRRR

            (R,K){20}
27: KKKKR KKKKKRKKKKRRRRRRRRRR RRRRR

            (R,K){20}
28: KKKRK KKKKRKKKKRRRRRRRRRRR RRRRR

            (R,K){20}
29: KKRKK KKKRKKKKRRRRRRRRRRRR RRRRR

            (R,K){20}
30: KRKKK KKRKKKKRRRRRRRRRRRRR RRRRR

            (R,K){20}
31: RKKKK KRKKKKRRRRRRRRRRRRRR RRRRR

            (R,K){20}
32: KKKKK RKKKKRRRRRRRRRRRRRRR RRRRR

            (R,K){20}
33: KKKKR KKKKRRRRRRRRRRRRRRRR RRRRR

            (R,K){20}
34: KKKRK KKKRRRRRRRRRRRRRRRRR RRRRR

            (R,K){20}
35: KKRKK KKRRRRRRRRRRRRRRRRRR RRRRR

            (R,K){20}
36: KRKKK KRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
37: RKKKK RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
38: KKKKR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
39: KKKRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
40: KKRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
41: KRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

             (R){20}
42: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNK

             (R){20}
43: RRRRR RRRRRRRRRRRRRRRRRRRR RRNKQ

             (R){20}
44: RRRRR RRRRRRRRRRRRRRRRRRRR RNKQT

             (R){20}
45: RRRRR RRRRRRRRRRRRRRRRRRRR NKQTK

AAM02687  ck: 1334  len: 86    ! Aam02687 Peptide #1369 encoded by probe for

1          (R,K){20,20}
            (R,K){20}
57: EEEEG RRRKKKKRKRKKKKKKKKKK KKKKK

            (R,K){20}
58: EEEGR RRKKKKRKRKKKKKKKKKKK KKKKK

            (R,K){20}
59: EEGRR RKKKKRKRKKKKKKKKKKKK KKKKK

            (R,K){20}
60: EGRRR KKKKRKRKKKKKKKKKKKKK KKKKK

            (R,K){20}
61: GRRRK KKKRKRKKKKKKKKKKKKKK KKKKK

            (R,K){20}
62: RRRKK KKRKRKKKKKKKKKKKKKKK KKKKK
```

```
             (R,K){20}
63: RRKKK KRKRKKKKKKKKKKKKKKKK KKKK

             (R,K){20}
64: RKKKK RKRKKKKKKKKKKKKKKKKK KKK

             (R,K){20}
65: KKKKR KRKKKKKKKKKKKKKKKKKK KK

             (R,K){20}
66: KKKRK RKKKKKKKKKKKKKKKKKKK K

              (K){20}
67: KKRKR KKKKKKKKKKKKKKKKKKKK

AAM02768  ck: 9082  len: 167   ! Aam02768 Peptide #1450 encoded by probe for
1           (R,K){20,20}
             (R,K){20}
33: EEGRG RRRRRRRKRRRKRRRKRRKR RRGGG

             (R,K){20}
34: EGRGR RRRRRRKRRRKRRRKRRKRR RGGGR

             (R,K){20}
35: GRGRR RRRRRKRRRKRRRKRRKRRR GGGRR

AAM03278  ck: 2276  len: 89    ! Aam03278 Peptide #1960 encoded by probe for
1           (R,K){20,20}
             (R,K){20}
23: EEEEE KKKKKKRKRKKKKRKKKKKK EEEKK

              (K){20}
46: KKEEE KKKKKKKKKKKKKKKKKKKK KRKEE

              (K){20}
47: KEEEK KKKKKKKKKKKKKKKKKKKK RKEEE

             (R,K){20}
48: EEEKK KKKKKKKKKKKKKKKKKKKR KEEEE

             (R,K){20}
49: EEKKK KKKKKKKKKKKKKKKKKKRK EEEEE

AAM03564  ck: 1939  len: 130   ! Aam03564 Peptide #2246 encoded by probe for
1           (R,K){20,20}
             (R,K){20}
42: EGRKE RRRRRRRRRRRRRRRRRKRK RRRRR

             (R,K){20}
43: GRKER RRRRRRRRRRRRRRRRKRKR RRRRR

             (R,K){20}
44: RKERR RRRRRRRRRRRRRRRKRKRR RRRRK

             (R,K){20}
45: KERRR RRRRRRRRRRRRRRKRKRRR RRRKK

             (R,K){20}
46: ERRRR RRRRRRRRRRRRRKRKRRRR RRKKK

             (R,K){20}
47: RRRRR RRRRRRRRRRRRKRKRRRRR RKKKK

             (R,K){20}
48: RRRRR RRRRRRRRRRRKRKRRRRRR KKKKK

             (R,K){20}
49: RRRRR RRRRRRRRRRKRKRRRRRRK KKKKK

             (R,K){20}
50: RRRRR RRRRRRRRRKRKRRRRRRKK KKKKE

             (R,K){20}
51: RRRRR RRRRRRRRKRKRRRRRRKKK KKKEE

             (R,K){20}
52: RRRRR RRRRRRRKRKRRRRRRKKKK KKEEE

             (R,K){20}
53: RRRRR RRRRRRKRKRRRRRRKKKKK KEEEE

             (R,K){20}
54: RRRRR RRRRRKRKRRRRRRKKKKKK EEEEE

AAM03848  ck: 3607  len: 88    ! Aam03848 Peptide #2530 encoded by probe for
1           (R,K){20,20}
              (K){20}
39: ERKRE KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
40: RKREK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
41: KREKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
42: REKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
43: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
```

```
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

           (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

           (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKK KK

           (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKK K

           (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKK

AAM04408  ck: 3937  len: 85   ! Aam04408 Peptide #3090 encoded by probe for

1         (R,K){20,20}
           (K){20}
 1:       KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

           (K){20}
 7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKE

           (K){20}
 8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKEE

           (K){20}
 9: KKKKK KKKKKKKKKKKKKKKKKKKK KKEEE

           (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KEEEE

           (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK EEEEX

AAM06100  ck: 1560  len: 88   ! Aam06100 Peptide #4782 encoded by probe for

1         (R,K){20,20}
           (R){20}
43: RERRG RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
44: ERRGR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
45: RRGRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
46: RGRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
47: GRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
48: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
49: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
50: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
51: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
52: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
53: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
54: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
55: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
56: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
57: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

           (R){20}
58: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRN

           (R){20}
59: RRRRR RRRRRRRRRRRRRRRRRRRR RRRNT

           (R){20}
60: RRRRR RRRRRRRRRRRRRRRRRRRR RRNTN

           (R){20}
61: RRRRR RRRRRRRRRRRRRRRRRRRR RNTNN

           (R){20}
62: RRRRR RRRRRRRRRRRRRRRRRRRR NTNNE

AAG73687  ck: 3063  len: 29   ! Aag73687 Human colon cancer antigen protein

1         (R,K){20,20}
```

```
              (K){20}
 8: MMMTF KKKKKKKKKKKKKKKKKKKK KX

              (K){20}
 9: MMTFK KKKKKKKKKKKKKKKKKKKK X

AAG73729  ck: 783  len: 83   ! Aag73729 Human colon cancer antigen protein
1
            (R,K){20,20}
              (K){20}
49: LGPCE KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
50: GPCEK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
51: PCEKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
52: CEKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
53: EKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGR

              (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KGGRX

              (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK GGRXK

AAG73810  ck: 3374  len: 88   ! Aag73810 Human colon cancer antigen protein
1
            (R,K){20,20}
              (K){20}
44: FGQTX KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
45: GQTXK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
46: QTXKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
47: TXKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
48: XKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKK KKGGP

              (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKK KGGPX

              (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKK GGPX

AAG73895  ck: 1887  len: 43   ! Aag73895 Human colon cancer antigen protein
1
            (R,K){20,20}
             (R,K){20}
18: VRPRV RKKKKKKKKKKKKKKKKKKK KKKKG

              (K){20}
19: RPRVR KKKKKKKKKKKKKKKKKKKK KKKGG

              (K){20}
20: PRVRK KKKKKKKKKKKKKKKKKKKK KKGG

              (K){20}
21: RVRKK KKKKKKKKKKKKKKKKKKKK KGG

              (K){20}
22: VRKKK KKKKKKKKKKKKKKKKKKKK GG

AAG74218  ck: 8659  len: 104   ! Aag74218 Human colon cancer antigen protein
1
            (R,K){20,20}
              (K){20}
75: PLGGQ KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
76: LGGQK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
77: GGQKK KKKKKKKKKKKKKKKKKKKK KKKXG

              (K){20}
78: GQKKK KKKKKKKKKKKKKKKKKKKK KKXGX

              (K){20}
79: QKKKK KKKKKKKKKKKKKKKKKKKK KXGXP

              (K){20}
80: KKKKK KKKKKKKKKKKKKKKKKKKK XGXPF
```

```
AAG74527   ck: 2664   len: 40    ! Aag74527 Human colon cancer antigen protein
1              (R,K){20,20}
                 (K){20}
  9: CLLLY KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 10: LLLYK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 11: LLYKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 12: LYKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 13: YKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                 (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGX

                 (K){20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKK KKGXP

                 (K){20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKK KGXPX

                 (K){20}
 17: KKKKK KKKKKKKKKKKKKKKKKKKK GXPX

AAG74650   ck: 1596   len: 69    ! Aag74650 Human colon cancer antigen protein
1              (R,K){20,20}
                 (K){20}
 36: LQCRQ KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 37: QCRQK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 38: CRQKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 39: RQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 40: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 41: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 42: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKG

                 (K){20}
 43: KKKKK KKKKKKKKKKKKKKKKKKKK KKKGX

                 (K){20}
 44: KKKKK KKKKKKKKKKKKKKKKKKKK KKGXP

                 (K){20}
 45: KKKKK KKKKKKKKKKKKKKKKKKKK KGXPF

                 (K){20}
 46: KKKKK KKKKKKKKKKKKKKKKKKKK GXPF

AAG74793   ck: 8497   len: 152   ! Aag74793 Human colon cancer antigen protein
1              (R,K){20,20}
                 (K){20}
122: SSHTQ KKKKKKKKKKKKKKKKKKKK KXKKK

                 (K){20}
123: SHTQK KKKKKKKKKKKKKKKKKKKK XKKKX

AAG74907   ck: 1215   len: 98    ! Aag74907 Human colon cancer antigen protein
1              (R,K){20,20}
                 (K){20}
 57: NLRKE KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 58: LRKEK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 59: RKEKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 60: KEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 61: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 62: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 63: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 64: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 65: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 66: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 67: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 68: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 69: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 70: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 71: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                 (K){20}
 72: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

                 (K){20}
 73: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXG

                 (K){20}
 74: KKKKK KKKKKKKKKKKKKKKKKKKK KKXGG

                 (K){20}
 75: KKKKK KKKKKKKKKKKKKKKKKKKK KXGG

                 (K){20}
 76: KKKKK KKKKKKKKKKKKKKKKKKKK XGG

AAG75215   ck: 3913   len: 155   ! Aag75215 Human colon cancer antigen protein
1              (R,K){20,20}
                 (K){20}
```

```
      135: RSSAP KKKKKKKKKKKKKKKKKKKK K

                   (K){20}
      136: SSAPK KKKKKKKKKKKKKKKKKKKK

       AAG75886   ck: 4235   len: 71    ! Aag75886 Human colon cancer antigen protein

1                 (R,K){20,20}
                   (K){20}
       47: KKKKX KKKKKKKKKKKKKKKKKKKK KXGXX

                   (K){20}
       48: KKKXK KKKKKKKKKKKKKKKKKKKK XGXX

       AAE01796   ck: 4416   len: 72    ! Aae01796 Human gene 27 encoded secreted pro

1                 (R,K){20,20}
                   (K){20}
       47: LPTFL KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       48: PTFLK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       49: TFLKK KKKKKKKKKKKKKKKKKKKK KKKK

                   (K){20}
       50: FLKKK KKKKKKKKKKKKKKKKKKKK KKK

                   (K){20}
       51: LKKKK KKKKKKKKKKKKKKKKKKKK KK

                   (K){20}
       52: KKKKK KKKKKKKKKKKKKKKKKKKK K

                   (K){20}
       53: KKKKK KKKKKKKKKKKKKKKKKKKK

       AAE01848   ck: 5584   len: 73    ! Aae01848 Human gene 27 encoded secreted pro

1                 (R,K){20,20}
                   (K){20}
       47: LPTFL KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       48: PTFLK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       49: TFLKK KKKKKKKKKKKKKKKKKKKK KKKKI

                   (K){20}
       50: FLKKK KKKKKKKKKKKKKKKKKKKK KKKI

                   (K){20}
       51: LKKKK KKKKKKKKKKKKKKKKKKKK KKI

                   (K){20}
       52: KKKKK KKKKKKKKKKKKKKKKKKKK KI

                   (K){20}
       53: KKKKK KKKKKKKKKKKKKKKKKKKK I

       AAB90574   ck: 1431   len: 530   ! Aab90574 Human secreted protein, SEQ ID NO:

1                 (R,K){20,20}
                   (K){20}
      511: LHAPP KKKKKKKKKKKKKKKKKKKK
```

```
       AAB45846   ck: 8137   len: 45    ! Aab45846 Nucleic acid transporter system pep

1                 (R,K){20,20}
                   (K){20}
        4:   YKA KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        5:  YKAK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        6: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        7: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        8: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
        9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                   (K){20}
       20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

                   (K){20}
       21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

                   (K){20}
       22: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

                   (K){20}
       23: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

                   (K){20}
       24: KKKKK KKKKKKKKKKKKKKKKKKKK WK

       AAB45848   ck: 4361   len: 59    ! Aab45848 Nucleic acid transporter system pep

1                 (R,K){20,20}
                   (K){20}
       18: APYKA KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
               (K){20}
19: PYKAK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
20: YKAKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
21: KAKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
22: AKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKW

               (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKKWK

               (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KKWK

               (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK KWK

               (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKK WK

AAB45850  ck: 4925  len: 100   ! Aab45850 Nucleic acid transporter system pe

1           (R,K){20,20}
               (K){20}
 1:           KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 2:         K KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 3:        KK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 4:       KKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 5:      KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 6:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 7:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 8:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
 9:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
10:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
11:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
12:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
13:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
14:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
15:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
16:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
17:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
18:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
19:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
20:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
21:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
22:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
23:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
24:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
25:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
26:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
27:     KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

               (K){20}
```

```
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
38: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
39: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
40: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
41: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
42: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
43: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
44: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
45: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
46: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
47: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
48: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
49: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
50: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
51: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
52: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
53: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
54: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
55: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
56: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
57: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
58: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
59: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
60: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
61: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
62: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
63: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
64: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
65: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
66: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
67: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
68: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
69: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
70: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
71: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
72: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
73: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
74: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
75: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

          (K){20}
76: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
  77: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
  78: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
  79: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
  80: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
  81: KKKKK KKKKKKKKKKKKKKKKKKKK

  AAB50247  ck: 8085  len: 154   ! Aab50247 Human breast cancer associated B72

1            (R,K){20,20}
              (K){20}
 114: TQLRQ KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 115: QLRQK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 116: LRQKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 117: RQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 118: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 121: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 122: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 123: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 124: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
 125: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

              (K){20}
 126: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

              (K){20}
 127: KKKKK KKKKKKKKKKKKKKKKKKKK KXKKK

              (K){20}
 128: KKKKK KKKKKKKKKKKKKKKKKKKK XKKKK

  ABJ37738  ck: 8085  len: 154   ! Abj37738 Human tumour-related protein - SEQ

1            (R,K){20,20}
              (K){20}
 114: TQLRQ KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 115: QLRQK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 116: LRQKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 117: RQKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 118: QKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 119: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 120: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 121: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 122: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 123: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 124: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX

              (K){20}
 125: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXK

              (K){20}
 126: KKKKK KKKKKKKKKKKKKKKKKKKK KKXKK

              (K){20}
 127: KKKKK KKKKKKKKKKKKKKKKKKKK KXKKK

              (K){20}
 128: KKKKK KKKKKKKKKKKKKKKKKKKK XKKKK

  ABR00951  ck: 6732  len: 139   ! Abr00951 Human gene 5-encoded secreted prot

1            (R,K){20,20}
              (K){20}
  91: HFRPG KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  92: FRPGK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  93: RPGKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  94: PGKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  95: GKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
              (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
108: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
109: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
110: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
111: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
112: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
113: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
115: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
116: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
117: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
118: KKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKK

 ABR01117  ck: 9531  len: 66    ! Abr01117 Human gene 171-encoded secreted pr

1            (R,K){20,20}
              (K){20}
 41: MWTVX KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 42: WTVXK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 43: TVXKK KKKKKKKKKKKKKKKKKKKK KKKK

              (K){20}
 44: VXKKK KKKKKKKKKKKKKKKKKKKK KKK

              (K){20}
 45: XKKKK KKKKKKKKKKKKKKKKKKKK KK

              (K){20}
 46: KKKKK KKKKKKKKKKKKKKKKKKKK K

              (K){20}
 47: KKKKK KKKKKKKKKKKKKKKKKKKK

 AAE33919  ck: 185   len: 22    ! Aae33919 Secretion domain peptide #17. 5/20

1            (R,K){20,20}
              (R){20}
  1:       RRRRRRRRRRRRRRRRRRRR GC

 ABG73150  ck: 5536  len: 84    ! Abg73150 Single-chain antigen-binding (sFv)

1            (R,K){20,20}
              (K){20}
  1:       KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  2:     K KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  3:    KK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  4:   KKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  5:  KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  6: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  7: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  8: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
  9: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
 16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

              (K){20}
```

```
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
              (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX
              (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXX
              (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKXXX
              (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KXXXX
              (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK XXXXX

ABG73151   ck: 8024   len: 84    ! Abg73151 Single-chain antigen-binding (sFv)

1            (R,K){20,20}
              (R){20}
 1:          RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
 2:        R RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
 3:    RR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
 4:   RRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
 5:  RRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
 6: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
 7: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
 8: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
 9: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
10: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
11: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
12: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
13: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
14: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
15: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
16: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
17: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
18: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
19: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
20: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
21: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
22: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
23: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
24: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
26: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
              (R){20}
27: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
```

```
             (R){20}
28: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
29: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
30: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
31: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
32: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
33: RRRRR RRRRRRRRRRRRRRRRRRR RRRRX

             (R){20}
34: RRRRR RRRRRRRRRRRRRRRRRRR RRRXX

             (R){20}
35: RRRRR RRRRRRRRRRRRRRRRRRR RRXXX

             (R){20}
36: RRRRR RRRRRRRRRRRRRRRRRRR RXXXX

             (R){20}
37: RRRRR RRRRRRRRRRRRRRRRRRR XXXXX

ABG73152  ck: 9643  len: 83    ! Abg73152 Single-chain antigen-binding (sFv)

1           (R,K){20,20}
             (R,K){20}
 1:         RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 2:       R KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 3:      RK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 4:     RKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 5:    RKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 6:   RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 7:   KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 8:   RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 9:   KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
10:   RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
11:   KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
12:   RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
13:   KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
14: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
15: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
16: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
17: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
18: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
19: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
20: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
21: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
22: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
23: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
24: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
25: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
26: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
27: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
28: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
29: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
30: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
31: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
32: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
33: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKX

             (R,K){20}
34: RKRKR KRKRKRKRKRKRKRKRKRKR KRKXX

             (R,K){20}
35: KRKRK RKRKRKRKRKRKRKRKRKRK RKXXX

             (R,K){20}
36: RKRKR KRKRKRKRKRKRKRKRKRKR KXXXX

             (R,K){20}
37: KRKRK RKRKRKRKRKRKRKRKRKRK XXXXX
```

```
   ABG73153  ck: 6708  len: 84   ! Abg73153 Single-chain antigen-binding (sFv)

1               (R,K){20,20}
                  (R){20}
     1:         RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
     2:       R RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
     3:      RR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
     4:     RRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
     5:    RRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
     6:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
     7:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
     8:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
     9:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    10:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    11:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    12:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    13:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    14:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    15:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    16:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    17:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    18:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    19:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    20:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    21:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    22:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    23:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    24:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    25:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    26:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    27:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    28:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    29:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    30:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    31:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    32:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

                  (R){20}
    33:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRRX

                  (R){20}
    34:   RRRRR RRRRRRRRRRRRRRRRRRRR RRRXX

                  (R){20}
    35:   RRRRR RRRRRRRRRRRRRRRRRRRR RRXXX

                  (R){20}
    36:   RRRRR RRRRRRRRRRRRRRRRRRRR RXXXX

                  (R){20}
    37:   RRRRR RRRRRRRRRRRRRRRRRRRR XXXXX

   ABG73870  ck: 5536  len: 84   ! Abg73870 Single-chain antigen-binding (sFv)

1               (R,K){20,20}
                  (K){20}
     1:         KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     2:       K KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     3:      KK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     4:     KKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     5:    KKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     6:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     7:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     8:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

                  (K){20}
     9:   KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
          (K){20}
10: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
11: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
12: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
13: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
14: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
15: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
16: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
17: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
18: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
19: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
20: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
21: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
22: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
23: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
24: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
25: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
26: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
27: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
28: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
29: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
30: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
31: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
32: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
          (K){20}
33: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKX
          (K){20}
34: KKKKK KKKKKKKKKKKKKKKKKKKK KKKXX
          (K){20}
35: KKKKK KKKKKKKKKKKKKKKKKKKK KKXXX
          (K){20}
36: KKKKK KKKKKKKKKKKKKKKKKKKK KXXXX
          (K){20}
37: KKKKK KKKKKKKKKKKKKKKKKKKK XXXXX

ABG73871  ck: 8024  len: 84    ! Abg73871 Single-chain antigen-binding (sFv)

1         (R,K){20,20}
          (R){20}
 1:       RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
 2:     R RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
 3:    RR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
 4:   RRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
 5:  RRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
 6: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
 7: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
 8: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
 9: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
10: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
11: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
12: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
13: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
14: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
15: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
16: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
17: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
18: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
19: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR
          (R){20}
```

```
20: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
21: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
22: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
23: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
24: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
26: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
27: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
28: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
29: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
30: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
31: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
32: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

          (R){20}
33: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRX

          (R){20}
34: RRRRR RRRRRRRRRRRRRRRRRRRR RRRXX

          (R){20}
35: RRRRR RRRRRRRRRRRRRRRRRRRR RRXXX

          (R){20}
36: RRRRR RRRRRRRRRRRRRRRRRRRR RXXXX

          (R){20}
37: RRRRR RRRRRRRRRRRRRRRRRRRR XXXXX

ABG73872   ck: 9643   len: 83    ! Abg73872 Single-chain antigen-binding (sFv)

1           (R,K){20,20}
             (R,K){20}
 1:         RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 2:       R KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 3:      RK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 4:     RKR KRKRKRKRKRKRKRKRKRKR KRKRK

             (R,K){20}
 5:    RKRK RKRKRKRKRKRKRKRKRKRK RKRKR

             (R,K){20}
 6: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
 7: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
 8: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
 9: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
10: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
11: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
12: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
13: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
14: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
15: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
16: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
17: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
18: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
19: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
20: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
21: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
22: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
23: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
24: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
25: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
26: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
27: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
28: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

          (R,K){20}
29: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

          (R,K){20}
30: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK
```

```
           (R,K){20}
31: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKR

           (R,K){20}
32: RKRKR KRKRKRKRKRKRKRKRKRKR KRKRK

           (R,K){20}
33: KRKRK RKRKRKRKRKRKRKRKRKRK RKRKX

           (R,K){20}
34: RKRKR KRKRKRKRKRKRKRKRKRKR KRKXX

           (R,K){20}
35: KRKRK RKRKRKRKRKRKRKRKRKRK RKXXX

           (R,K){20}
36: RKRKR KRKRKRKRKRKRKRKRKRKR KXXXX

           (R,K){20}
37: KRKRK RKRKRKRKRKRKRKRKRKRK XXXXX

ABG73873  ck: 6708  len: 84   ! Abg73873 Single-chain antigen-binding (sFv)

1          (R,K){20,20}
             (R){20}
 1:       RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 2:     R RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 3:    RR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 4:   RRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 5:  RRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 6: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 7: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 8: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
 9: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
10: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
11: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
12: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
13: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
14: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
15: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
16: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
17: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
18: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
19: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
20: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
21: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
22: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
23: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
24: RRRRR RRRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
25: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
26: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
27: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
28: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
29: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
30: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
31: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
32: RRRRR RRRRRRRRRRRRRRRRRRR RRRRR

             (R){20}
33: RRRRR RRRRRRRRRRRRRRRRRRR RRRRX

             (R){20}
34: RRRRR RRRRRRRRRRRRRRRRRRR RRRXX

             (R){20}
35: RRRRR RRRRRRRRRRRRRRRRRRR RRXXX

             (R){20}
36: RRRRR RRRRRRRRRRRRRRRRRRR RXXXX

             (R){20}
37: RRRRR RRRRRRRRRRRRRRRRRRR XXXXX

ABJ26082  ck: 5116  len: 448   ! Abj26082 Aspergillus fumigatus essential ger

1          (R,K){20,20}
             (K){20}
293: EHRKG KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
294: HRKGK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
            (K){20}
295: RKGKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
296: KGKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
297: GKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
298: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
299: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
300: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
301: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
302: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKA

            (K){20}
303: KKKKK KKKKKKKKKKKKKKKKKKKK KKKAP

            (K){20}
304: KKKKK KKKKKKKKKKKKKKKKKKKK KKAPF

            (K){20}
305: KKKKK KKKKKKKKKKKKKKKKKKKK KAPFP

            (K){20}
306: KKKKK KKKKKKKKKKKKKKKKKKKK APFPL

            (K){20}
342: PYKGE KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
343: YKGEK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
344: KGEKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
345: GEKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
346: EKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
347: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
348: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
349: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
350: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKD

            (K){20}
351: KKKKK KKKKKKKKKKKKKKKKKKKK KKKDT

            (K){20}
352: KKKKK KKKKKKKKKKKKKKKKKKKK KKDTL

            (K){20}
353: KKKKK KKKKKKKKKKKKKKKKKKKK KDTLT

            (K){20}
354: KKKKK KKKKKKKKKKKKKKKKKKKK DTLTH

ABJ19654  ck: 6732  len: 139   ! Abj19654 Human secreted protein amino acid

1           (R,K){20,20}
            (K){20}
91: HFRPG KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
92: FRPGK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
93: RPGKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
94: PGKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
95: GKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
108: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
109: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
110: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
111: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

            (K){20}
112: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK
```

```
             (K){20}
113: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
115: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
116: KKKKK KKKKKKKKKKKKKKKKKKKKK KKKK

             (K){20}
117: KKKKK KKKKKKKKKKKKKKKKKKKKK KKK

             (K){20}
118: KKKKK KKKKKKKKKKKKKKKKKKKKK KK

             (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKKK K

             (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKKK

 ABP99475   ck: 6732  len: 139   ! Abp99475 Human secreted protein SEQ ID NO 4

1           (R,K){20,20}
             (K){20}
 91: HFRPG KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 92: FRPGK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 93: RPGKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 94: PGKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 95: GKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 96: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 97: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 98: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 99: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
100: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
101: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
102: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
103: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
104: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
105: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
106: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
107: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
108: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
109: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
110: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
111: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
112: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
113: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
114: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
115: KKKKK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
116: KKKKK KKKKKKKKKKKKKKKKKKKK KKKK

             (K){20}
117: KKKKK KKKKKKKKKKKKKKKKKKKK KKK

             (K){20}
118: KKKKK KKKKKKKKKKKKKKKKKKKK KK

             (K){20}
119: KKKKK KKKKKKKKKKKKKKKKKKKK K

             (K){20}
120: KKKKK KKKKKKKKKKKKKKKKKKKK

 ABP99639   ck: 9531  len: 66    ! Abp99639 Human secreted protein SEQ ID NO 5

1           (R,K){20,20}
             (K){20}
 41: MWTVX KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 42: WTVXK KKKKKKKKKKKKKKKKKKKK KKKKK

             (K){20}
 43: TVXKK KKKKKKKKKKKKKKKKKKKK KKKK

             (K){20}
 44: VXKKK KKKKKKKKKKKKKKKKKKKK KKK

             (K){20}
 45: XKKKK KKKKKKKKKKKKKKKKKKKK KK

             (K){20}
 46: KKKKK KKKKKKKKKKKKKKKKKKKK K

             (K){20}
 47: KKKKK KKKKKKKKKKKKKKKKKKKK

Databases searched:
      Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003
```

Total finds:	8,655
Total length:	158,726,570
Total sequences:	1,107,863
CPU time:	07:06.66

```
! FINDPATTERNS on geneseqp:* allowing 0 mismatches

!        1 C(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V)


            AAB13781  ck: 1303  len: 25    ! Aab13781 Soluble peptide antigen pEA. 11/20

1                    C(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I
                       C(E)(A){5}(E)(A){5}(E)(A){5}
             1:        CEAAAAAEAAAAAEAAAAA   EAAAA


            AAB13783  ck: 4553  len: 45    ! Aab13783 Soluble tandem pEA/ pK peptide con

1                    C(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I
                       C(E)(A){5}(E)(A){5}(E)(A){5}
             1:        CEAAAAAEAAAAAEAAAAA   EAAAA


 Databases searched:
        Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

    Total finds:              2
   Total length:    158,726,570
 Total sequences:      1,107,863
       CPU time:       05:59.54
```

```
!!SEQUENCE_LIST 1.0
! FINDPATTERNS on geneseqp:* allowing 0 mismatches

!       1 C(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V)

GENESEQP2000S:AAB13781  ck: 1303  len: 25     finds: 1    ! Aab13781 Soluble peptide ant

GENESEQP2000S:AAB13783  ck: 4553  len: 45     finds: 1    ! Aab13783 Soluble tandem pEA/

\\End of list

 Databases searched:
        Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

     Total finds:            2
    Total length:  158,726,570
 Total sequences:    1,107,863
        CPU time:     08:18.53
```

```
!!AA_SEQUENCE 1.0
ID   AAB13781 standard; peptide; 25 AA.
XX
AC   AAB13781;
XX
DT   10-NOV-2000  (first entry)
XX
DE   Soluble peptide antigen pEA.
XX
KW   pEA peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW   major histocompatibility complex class 1; MHC class 1; antigen; tumour;
KW   prostate; breast; multiple myeloma.
XX
OS   Unidentified.
XX
PN   WO200035949-A1.
XX
PD   22-JUN-2000.
XX
PF   14-DEC-1999;   99WO-US29724.
XX
PR   14-DEC-1998;   98US-0112324.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Hakim I,  Vidovic D;
XX
DR   WPI; 2000-442365/38.
XX
PT   Antigens modified by the covalent addition of a peptide that
PT   facilitates entry into antigen presenting cells, useful for producing
PT   compositions for immunizing against tumors and pathogens -
XX
PS   Claim 2; Page 26; 34pp; English.
XX
CC   The present invention relates to compositions of modified soluble protein
CC   antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC   response i.e. a major histocompatability complex (MHC) class I molecule
CC   response. The protein antigen is modified by the covalent addition of a
CC   peptide sequence which facilitate entry of the antigen into antigen
CC   presenting cells (APCs). The present sequence is one such peptide
CC   sequence which can be used to modify the soluble antigens. The present
CC   sequence is peptide pEA. The modified antigen composition may be used for
CC   immunising against, or treating a tumour e.g. prostate and breast
CC   carcinoma or multiple myeloma, or pathogen in mammals.
XX
SQ   Sequence   25 AA;

 AAB13781  Length: 25  January 30, 2004 10:59  Type: P  Check: 1303  ..

       1  CEAAAAAEAA AAAEAAAAAE AAAAA

!!AA_SEQUENCE 1.0
ID   AAB13783 standard; peptide; 45 AA.
XX
AC   AAB13783;
XX
DT   10-NOV-2000  (first entry)
XX
DE   Soluble tandem pEA/ pK peptide conjugate.
XX
KW   pk peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW   major histocompatibility complex class 1; MHC class 1; antigen; tumour;
KW   prostate; breast; multiple myeloma; pEA peptide.
XX
OS   Unidentified.
XX
PN   WO200035949-A1.
XX
PD   22-JUN-2000.
XX
PF   14-DEC-1999;   99WO-US29724.
XX
PR   14-DEC-1998;   98US-0112324.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Hakim I,  Vidovic D;
XX
DR   WPI; 2000-442365/38.
XX
PT   Antigens modified by the covalent addition of a peptide that
PT   facilitates entry into antigen presenting cells, useful for producing
PT   compositions for immunizing against tumors and pathogens -
XX
PS   Claim 2; Page 26; 34pp; English.
XX
CC   The present invention relates to compositions of modified soluble protein
CC   antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC   response i.e. a major histocompatability complex (MHC) class I molecule
CC   response. The protein antigen is modified by the covalent addition of a
CC   peptide sequence which facilitate entry of the antigen into antigen
CC   presenting cells (APCs). The present sequence is one such peptide
CC   sequence which can be used to modify the soluble antigens. The present
CC   sequence is tandem pEA/ pK peptide conjugate. The modified antigen
CC   composition may be used for immunising against, or treating a tumour e.g.
CC   prostate and breast carcinoma or multiple myeloma, or pathogen in
CC   mammals.
XX
SQ   Sequence   45 AA;

 AAB13783  Length: 45  January 30, 2004 10:59  Type: P  Check: 4553  ..

       1  CEAAAAAEAA AAAEAAAAAE AAAAAKKKKK KKKKKKKKKK KKKKK
```

```
! FINDPATTERNS on pir:* allowing 0 mismatches

!        1 C(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V)

 Databases searched:
        NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

     Total finds:             0
    Total length:    96,168,682
 Total sequences:       283,308
        CPU time:      02:26.82
```

```
! FINDPATTERNS on swp:* allowing 0 mismatches

!        1 C(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V)


Databases searched:
       SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
       SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

    Total finds:            0
   Total length:  305,079,309
Total sequences:      958,388
       CPU time:     08:14.71
```

```
! FINDPATTERNS on geneseqp:* allowing 0 mismatches

!        1 (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){


            AAB13781  ck: 1303  len: 25    ! Aab13781 Soluble peptide antigen pEA. 11/20

1                     (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                        (E)(A){5}(E)(A){5}(E)(A){5}
             2:     C   EAAAAAEAAAAAEAAAAA   EAAAA

                        (E)(A){5}(E)(A){5}(E)(A){5}
             8: AAAAA   EAAAAAEAAAAAEAAAAA


            AAB13783  ck: 4553  len: 45    ! Aab13783 Soluble tandem pEA/ pK peptide con

1                     (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                        (E)(A){5}(E)(A){5}(E)(A){5}
             2:     C   EAAAAAEAAAAAEAAAAA   EAAAA

                        (E)(A){5}(E)(A){5}(E)(A){5}
             8: AAAAA   EAAAAAEAAAAAEAAAAA   KKKKK


            ABP02760  ck: 5947  len: 86    ! Abp02760 Human ORFX protein sequence SEQ ID

1                     (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                        (D)(A,G,C){5}(E)(A,L,G,C,V){5}(E)(A,G,C,V){5}
            13: GGRVR   DACGCCEVCGALEGAVCG   LQEGP

Databases searched:
      Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

    Total finds:            5
   Total length:  158,726,570
Total sequences:    1,107,863
       CPU time:     09:16.70
```

```
!!SEQUENCE_LIST 1.0
! FINDPATTERNS on geneseqp:* allowing 0 mismatches

!       1 (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){

GENESEQP2000S:AAB13781  ck: 1303  len: 25    finds: 2    ! Aab13781 Soluble peptide ant

GENESEQP2000S:AAB13783  ck: 4553  len: 45    finds: 2    ! Aab13783 Soluble tandem pEA/

GENESEQP2002S:ABP02760  ck: 5947  len: 86    finds: 1    ! Abp02760 Human ORFX protein

\\End of list

 Databases searched:
        Geneseq-AA, Release 13.0, Released on 19Jun2003, Formatted on 15Jul2003

     Total finds:              5
    Total length:    158,726,570
 Total sequences:      1,107,863
        CPU time:       11:32.04
```

```
!!AA_SEQUENCE 1.0
ID   AAB13781 standard; peptide; 25 AA.
XX
AC   AAB13781;
XX
DT   10-NOV-2000  (first entry)
XX
DE   Soluble peptide antigen pEA.
XX
KW   pEA peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW   major histocompatibility complex class 1; MHC class 1; antigen; tumour;
KW   prostate; breast; multiple myeloma.
XX
OS   Unidentified.
XX
PN   WO200035949-A1.
XX
PD   22-JUN-2000.
XX
PF   14-DEC-1999;   99WO-US29724.
XX
PR   14-DEC-1998;   98US-0112324.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Hakim I,  Vidovic D;
XX
DR   WPI; 2000-442365/38.
XX
PT   Antigens modified by the covalent addition of a peptide that
PT   facilitates entry into antigen presenting cells, useful for producing
PT   compositions for immunizing against tumors and pathogens -
XX
PS   Claim 2; Page 26; 34pp; English.
XX
CC   The present invention relates to compositions of modified soluble protein
CC   antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC   response i.e. a major histocompatability complex (MHC) class I molecule
CC   response. The protein antigen is modified by the covalent addition of a
CC   peptide sequence which facilitate entry of the antigen into antigen
CC   presenting cells (APCs). The present sequence is one such peptide
CC   sequence which can be used to modify the soluble antigens. The present
CC   sequence is peptide pEA. The modified antigen composition may be used for
CC   immunising against, or treating a tumour e.g. prostate and breast
CC   carcinoma or multiple myeloma, or pathogen in mammals.
XX
SQ   Sequence   25 AA;

 AAB13781  Length: 25  January 30, 2004 11:00  Type: P  Check: 1303  ..

       1  CEAAAAAEAA AAAEAAAAAE AAAAA

!!AA_SEQUENCE 1.0
ID   AAB13783 standard; peptide; 45 AA.
XX
AC   AAB13783;
XX
DT   10-NOV-2000  (first entry)
XX
DE   Soluble tandem pEA/ pK peptide conjugate.
XX
KW   pk peptide; cytostatic; vaccine; cytotoxic T cell; CTL; immunotherapy;
KW   major histocompatibility complex class 1; MHC class 1; antigen; tumour;
KW   prostate; breast; multiple myeloma; pEA peptide.
XX
OS   Unidentified.
XX
PN   WO200035949-A1.
XX
PD   22-JUN-2000.
XX
PF   14-DEC-1999;   99WO-US29724.
XX
PR   14-DEC-1998;   98US-0112324.
XX
PA   (DEND-) DENDREON CORP.
XX
PI   Laus R,  Hakim I,  Vidovic D;
XX
DR   WPI; 2000-442365/38.
XX
PT   Antigens modified by the covalent addition of a peptide that
PT   facilitates entry into antigen presenting cells, useful for producing
PT   compositions for immunizing against tumors and pathogens -
XX
PS   Claim 2; Page 26; 34pp; English.
XX
CC   The present invention relates to compositions of modified soluble protein
CC   antigens capable of eliciting an enhanced in vivo cytotoxic T cell (CTL)
CC   response i.e. a major histocompatability complex (MHC) class I molecule
CC   response. The protein antigen is modified by the covalent addition of a
CC   peptide sequence which facilitate entry of the antigen into antigen
CC   presenting cells (APCs). The present sequence is one such peptide
CC   sequence which can be used to modify the soluble antigens. The present
CC   sequence is tandem pEA/ pK peptide conjugate. The modified antigen
CC   composition may be used for immunising against, or treating a tumour e.g.
CC   prostate and breast carcinoma or multiple myeloma, or pathogen in
CC   mammals.
XX
SQ   Sequence   45 AA;

 AAB13783  Length: 45  January 30, 2004 11:00  Type: P  Check: 4553  ..

       1  CEAAAAAEAA AAAEAAAAAE AAAAAKKKKK KKKKKKKKKK KKKKK

!!AA_SEQUENCE 1.0
ID   ABP02760 standard; Protein; 86 AA.
XX
AC   ABP02760;
XX
DT   25-JUN-2002  (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:5502.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   N-PSDB; ABN18512.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders  -
XX
PS   Disclosure; SEQ ID 5502; 1037pp; English.
```

```
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   86 AA;

 ABP02760  Length: 86  January 30, 2004 11:00  Type: P  Check: 5947  ..

       1  XPPTDCEGGR VRDACGCCEV CGALEGAVCG LQEGPCGEGA ANAVSAPPSG

      51  VPASATVRRR AQAGLCVCAS SEPVCGNDAK TYTNLC
```

```
! FINDPATTERNS on swp:* allowing 0 mismatches

!        1 (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){


         HRA1_MOUSE  ck: 7206  len: 480   ! Q9r118 mus musculus (mouse). serine proteas

1                    (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                       (D)(A,G,C){5}(E)(A,L,G,C,V){5}(E)(A,G,C){5}
           60: GGRVR   DACGCCEVCGALEGAACG    LQEGP


             Q9KHD9  ck: 1697  len: 262   ! Q9khd9 streptomyces griseus subsp. griseus.

1                    (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                       (E)(A,L,C){5}(D)(A,L,G,V){5}(E)(A,L,F,G){5}
          124: LGGGL   ELALACDLVVAGEGALFG    LPELG


             Q8RUS6  ck: 3612  len: 107   ! Q8rus6 oryza sativa (japonica cultivar-grou

1                    (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                       (E)(A,I,V){5}(E)(G,C,V){5}(D)(A,L,G,M,V){5}
           20: LETEV   EAAIAVEGGCGVDLAMVG    RALGL


             Q9QZK6  ck: 7594  len: 480   ! Q9qzk6 mus musculus (mouse). insulin-like g

1                    (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                       (D)(A,G,C){5}(E)(A,L,G,C,V){5}(E)(A,G,C){5}
           60: GGRVR   DACGCCEVCGALEGAACG    LQEGP


             Q91WS3  ck: 7882  len: 480   ! Q91ws3 mus musculus (mouse). protease, seri

1                    (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                       (D)(A,G,C){5}(E)(A,L,G,C,V){5}(E)(A,G,C){5}
           60: GGRVR   DACGCCEVCGALEGAACG    LQEGP


             Q9QZK5  ck: 8689  len: 480   ! Q9qzk5 rattus norvegicus (rat). insulin-lik

1                    (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,
                       (D)(A,G,C){5}(E)(A,L,G,C,V){5}(E)(A,G,C,V){5}
           60: GGRVR   DACGCCEVCGALEGAVCG    LQEGP

 Databases searched:
        SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
        SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

      Total finds:          6
     Total length: 305,079,309
  Total sequences:    958,388
         CPU time:   14:58.50
```

```
!!SEQUENCE_LIST 1.0
! FINDPATTERNS on swp:* allowing 0 mismatches

!        1 (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){

SW:HRA1_MOUSE        ck: 7206  len: 480   finds: 1   ! Q9r118 mus musculus (mouse). s
SP_BA:Q9KHD9         ck: 1697  len: 262   finds: 1   ! Q9khd9 streptomyces griseus su
SP_PL:Q8RUS6         ck: 3612  len: 107   finds: 1   ! Q8rus6 oryza sativa (japonica
SP_RO:Q9QZK6         ck: 7594  len: 480   finds: 1   ! Q9qzk6 mus musculus (mouse). i
SP_RO:Q91WS3         ck: 7882  len: 480   finds: 1   ! Q91ws3 mus musculus (mouse). p
SP_RO:Q9QZK5         ck: 8689  len: 480   finds: 1   ! Q9qzk5 rattus norvegicus (rat)
\\End of list

 Databases searched:
        SWISS-PROT, Release 41.1, Released on 6Jun2003, Formatted on 9Jun2003
        SPTREMBL, Release 23.0, Released on 4Mar2003, Formatted on 7Mar2003

     Total finds:            6
    Total length:  305,079,309
 Total sequences:      958,388
        CPU time:     16:49.45
```

```
!!AA_SEQUENCE 1.0
ID   HRA1_MOUSE     STANDARD;      PRT;   480 AA.
AC   Q9R118;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Serine protease HTRA1 precursor (EC 3.4.21.-).
GN   PRSS11 OR HTRA1 OR HTRA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=Brain;
RA   Oka C., Soma A., Kanda H., Kawaichi M.;
RT   "The role of murine serine protease HTRA in osteogenesis.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Protease that regulate the availability of IGFs by
CC       cleaving IGF-binding proteins (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC   -!- SIMILARITY: Contains 1 IGFBP domain.
CC   -!- SIMILARITY: Contains 1 Kazal-like domain.
CC   -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF172994; AAD49422.1; -.
DR   MEROPS; S01.277; -.
DR   MGD; MGI:1929076; Prss11.
DR   InterPro; IPR000867; Insl_gro_fac_pr.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001940; Protease2C.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00050; kazal; 1.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00834; PROTEASES2C.
DR   SMART; SM00121; IB; 1.
DR   SMART; SM00280; KAZAL; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Hydrolase; Serine protease; Growth factor binding; Signal.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    480       SERINE PROTEASE HTRA1.
FT   DOMAIN       37     94       IGFBP.
FT   DOMAIN      101    155       KAZAL-LIKE.
FT   DOMAIN      204    364       SERINE PROTEASE.
FT   DOMAIN      365    467       PDZ.
FT   ACT_SITE    220    220       CHARGE RELAY SYSTEM (POTENTIAL).
FT   ACT_SITE    250    250       CHARGE RELAY SYSTEM (POTENTIAL).
FT   ACT_SITE    328    328       CHARGE RELAY SYSTEM (POTENTIAL).
SQ   SEQUENCE   480 AA;  51246 MW;  54BB9BA6C99A7BF4 CRC64;

HRA1_MOUSE  Length: 480  January 30, 2004 15:08  Type: P  Check: 7206  ..

       1  MQSLRTTLLS LLLLLLAAPS LALPSGTGRS APAATVCPEH CDPTRCAPPP

      51  TDCEGGRVRD ACGCCEVCGA LEGAACGLQE GPCGEGLQCV LPFGVPASAT

     101  VRRRAQAGLC VCASSEPVCG SDAKTYTNLC QLRAASRRSE KLPQPPVIVL

     151  QRGACGQGQE DPNSLRHKYN FIADVVEKFA PDVVKHELYR KLPFSKREVP

     201  VASGSGFIVS EDGLIVTNAH VVTNKNRVKV ELKNGATYEA IIKDVDEKAD

     251  IALIKIDHQG KLPVLLLGRS SELRPGEFVV AIGSPFSLQN TVTTGIVSTT

     301  QRGGKELGLR NSDMDYIQTD AIINYGNSGG PLVNLDGEVI GINTLKVTAG

     351  ISFAIPSDKI KKFLTESHDR QAKGKAVTKK KYIGIRMMSL TSSKAKELKD

     401  RHRDFPDVLS GAYIIEVIPD TPAEAGGLKE NDVIISINGQ SVVTANDVSD

     451  VIKKENTLNM VVRRGNEDIV ITVIPEEIDP

!!AA_SEQUENCE 1.0
ID   Q9KHD9     PRELIMINARY;      PRT;   262 AA.
AC   Q9KHD9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Enoyl-CoA hydratase-like protein.
OS   Streptomyces griseus subsp. griseus.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=67263;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM40695;
RX   MEDLINE=20316791; PubMed=10858335;
RA   Smith W.C., Xiang L., Shen B.;
RT   "Genetic localization and molecular characterization of the nonS gene
RT   required for macrotetrolide biosynthesis in Streptomyces griseus
RT   DSM40695.";
RL   Antimicrob. Agents Chemother. 44:1809-1817(2000).
DR   EMBL; AF263011; AAF81231.1; -.
DR   HSSP; P14604; 2DUB.
DR   InterPro; IPR001753; EnCoA_hydrtse.
DR   Pfam; PF00378; ECH; 1.
DR   PROSITE; PS00166; ENOYL_COA_HYDRATASE; 1.
SQ   SEQUENCE   262 AA;  27068 MW;  D218FA2BDEC64BF2 CRC64;

 Q9KHD9  Length: 262  January 30, 2004 15:08  Type: P  Check: 1697  ..

       1  MTPQETAPEG EPSPLLVERH GRVATLTLNR PHRRNAMSTA MLARLDHALG

      51  KLTDQGGEAP GALVLTGAGG TFSSGADTRE PDWRDLSRRA VRRAHFRTVF

     101  AMLHEAPFPV VAAVEGYALG GGLELALACD LVVAGEGALF GLPELGVGAV

     151  PGGGAVHSLV RRAGRGVAAR MLLIPGERVR ADELARLGAV ERTVPDGGAL

     201  AEAQALAASV AAGDPALLEA GVRLLRDSGH LDRTAALGVE NGYWWQAASA

     251  ANRDPSPSSG RP

!!AA_SEQUENCE 1.0
ID   Q8RUS6     PRELIMINARY;      PRT;   107 AA.
AC   Q8RUS6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   P0678F11.3 protein (P0413C03.30 protein).
GN   P0678F11.3 OR P0413C03.30.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0678F11.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0413C03.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003437; BAB86096.1; -.
DR   EMBL; AP003451; BAB86154.1; -.
DR   Gramene; Q8RUS6; -.
SQ   SEQUENCE   107 AA;  11046 MW;  89097D882018D849 CRC64;

 Q8RUS6  Length: 107  January 30, 2004 15:08  Type: P  Check: 3612  ..

       1  MPSPRRAVNL RWSSLETEVE AAIAVEGGCG VDLAMVGRAL GLDPATVRLN

      51  GYFVSRGRGH VSSAVTWRAL LDFFAARGLP TGDAPAAPVA VHGKPAPPPP

     101  PPPVSVL

!!AA_SEQUENCE 1.0
ID   Q9QZK6      PRELIMINARY;      PRT;   480 AA.
AC   Q9QZK6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Insulin-like growth factor binding protein 5 protease.
GN   PRSS11.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Ovary;
RA   Hourvitz A., Hennebold J.D., King G., Negishi H., Erickson G.F.,
RA   Roby J.A., Mayo K.E., Adashi E.Y.;
RT   "Mouse insulin-like growth factor binding protein 5-directed
RT   endopeptidase: structural assessment, evolutionary analysis, ovarian
RT   expression, hormonal regulation and cellular localization.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AF179369; AAD52682.1; -.
DR   MGD; MGI:1929076; Prss11.
DR   InterPro; IPR000867; Insl_gro_fac_pr.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001940; Protease2C.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00219; IGFBP; 1.
DR   Pfam; PF00050; kazal; 1.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00834; PROTEASES2C.
DR   SMART; SM00121; IB; 1.
DR   SMART; SM00280; KAZAL; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Hydrolase; Protease; Serine protease.
SQ   SEQUENCE   480 AA;  51213 MW;  92BDDA85CF5B12B7 CRC64;

 Q9QZK6  Length: 480  January 30, 2004 15:08  Type: P  Check: 7594  ..

       1  MQSLRTTLLS LLLLLLAAPS LALPSGTGRS APAATVCPEH CDPTRCAPPP

      51  TDCEGGRVRD ACGCCEVCGA LEGAACGLQE GPCGEGLQCV VPFGVPASAT

     101  VRRRAQAGLC VCASSEPVCG SDAKTYTNLC QLRAASRRSE KLRQPPVIVL

     151  QRGACGQGQE DPNSLRHKYN FIADVVEKIA PAVVHIELYR KLPFSKREVP

     201  VASGSGFIVS EDGLIVTNAH VVTNKNRVKV ELKNGATYEA KIKDVDEKAD

     251  IALIKIDHQG KLPVLLLGRS SELRPGEFVV AIGSPFSLQN TVTTGIVSTT

     301  QRGGKELGLR NSDMDYIQTD AIINYGNSGG PLVNLDGEVI GINTLKVTAG

     351  ISFAIPSDKI KKFLTESHDR QAKGKAVTKK KYIGIRMMSL TSSKAKELKD

     401  RHRDFPDVLS GAYIIEVIPD TPAEAGGLKE NDVIISINGQ SVVTANDVSD

     451  VIKKENTLNM VVRRGNEDIV ITVIPEEIDP

!!AA_SEQUENCE 1.0
ID   Q91WS3      PRELIMINARY;      PRT;   480 AA.
AC   Q91WS3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Protease, serine, 11 (Igf binding).
GN   PRSS11.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; BC013516; AAH13516.1; -.
DR   MGD; MGI:1929076; Prss11.
DR   InterPro; IPR000867; Insl_gro_fac_pr.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001940; Protease2C.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00219; IGFBP; 1.
DR   Pfam; PF00050; kazal; 1.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00834; PROTEASES2C.
DR   SMART; SM00121; IB; 1.
DR   SMART; SM00280; KAZAL; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Hydrolase; Protease; Serine protease.
SQ   SEQUENCE   480 AA;  51212 MW;  76BDD5E862EDC9DA CRC64;

 Q91WS3  Length: 480  January 30, 2004 15:08  Type: P  Check: 7882  ..

       1  MQSLRTTLLS LLLLLLAAPS LALPSGTGRS APAATVCPEH CDPTRCAPPP

      51  TDCEGGRVRD ACGCCEVCGA LEGAACGLQE GPCGEGLQCV VPFGVPASAT

     101  VRRRAQAGLC VCASSEPVCG SDAKTYTNLC QLRAASRRSE KLRQPPVIVL

     151  QRGACGQGQE DPNSLRHKYN FIADVVEKIA PAVVHIELYR KLPFSKREVP

     201  VASGSGFIVS EDGLIVTNAH VVTNKNRVKV ELKNGATYEA KIKDVDEKAD

     251  IALIKIDHQG KLPVLLLGRS SELRPGEFVV AIGSPFSLQN TVTTGIVSTT

     301  QRGGKELGLR NSDMDYIQTD AIINYGNSGG PLVNLDGEVI GINTLKVTAG

     351  ISFAIPSDKI KKFLTQSHDR QAKGKAVTKK KYIGIRMMSL TSSKAKELKD

     401  RHRDFPDVLS GAYIIEVIPD TPAEAGGLKE NDVIISINGQ SVVTANDVSD

     451  VIKKENTLNM VVRRGNEDIV ITVIPEEIDP

!!AA_SEQUENCE 1.0
ID   Q9QZK5      PRELIMINARY;      PRT;   480 AA.
AC   Q9QZK5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Insulin-like growth factor binding protein 5 protease.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Ovary;
RA   Hourvitz A., Hennebold J.D., King G., Negishi H., Erickson G.F.,
RA   Roby J.A., Mayo K.E., Adashi E.Y.;
RT   "Mouse insulin-like growth factor binding protein 5-directed
RT   endopeptidase: structural assessment, evolutionary analysis, ovarian
RT   expression, hormonal regulation and cellular localization.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR   EMBL; AF179370; AAD52683.1; -.
DR   MEROPS; S01.277; -.
DR   InterPro; IPR000867; Insl_gro_fac_pr.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001940; Protease2C.
DR   InterPro; IPR001254; Ser_protease_Try.
DR   Pfam; PF00219; IGFBP; 1.
DR   Pfam; PF00050; kazal; 1.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00834; PROTEASES2C.
DR   SMART; SM00121; IB; 1.
DR   SMART; SM00280; KAZAL; 1.
DR   SMART; SM00228; PDZ; 1.
DR   PROSITE; PS50106; PDZ; 1.
KW   Hydrolase; Protease; Serine protease.
SQ   SEQUENCE   480 AA;  51330 MW;  37A864C5A8FFC035 CRC64;

 Q9QZK5  Length: 480  January 30, 2004 15:08  Type: P  Check: 8689  ..

       1  MQFLRTALLS LLLLLLAAPS LALPSGISRS APAATVCPEH CDPTRCAPPP

      51  TDCEGGRVRD ACGCCEVCGA LEGAVCGLQE GPCGEGLQCV VPFGVPASAT

     101  VRRRAQAGLC VCASSEPVCG SDAKTYTNLC QLRAASRRSE KLRQPPVIVL

     151  QRGACGQGQE DPNSLRHKYN FIADVVEKIA PAVVHIELYR KLPFSKREVP

     201  VASGSGFIVS EDGLIVTNAH VVTNKNRVKV ELKNGATYEA KIKDVDEKAD

     251  IALIKIDHQG KLPVLLLGRS SELRPGEFVV AIGSPFSLQN TVTTGIVSTT

     301  QRGGKELGLR NSDMDYIQTD AIINYGNSGG PLVNLDGEVI GINTLKVTAG

     351  ISFAIPSDKI KKFLTESHDR QAKGKTVTKK KYIGIRMMSL TSSKAKELKD

     401  RHRDFPDVIS GAYIIEVIPD TPAEAGGLKE NDVIISINGQ SVVTANDVSD

     451  VIKKENTLNM VVRRGNEDIV ITVVPEEIDP
```

```
! FINDPATTERNS on pir:* allowing 0 mismatches

!        1 (E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){5,5}(E,D)(A,L,I,F,G,C,M,V){

 Databases searched:
        NBRF, Release 76.1, Released on 12May2003, Formatted on 10Jun2003

     Total finds:             0
    Total length:    96,168,682
 Total sequences:       283,308
        CPU time:      04:36.11
```